2006

The Relationship Between the Virginia Standards of Learning Tests and the New PSAT/NMSQT

Susan P. McKelvey
Virginia Commonwealth University

The Relationship Between

the Virginia Standards of Learning Tests

and the New PSAT/NMSQT

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

By
Susan P. McKelvey
BA, Randolph-Macon College, French
M.Ed., Virginia Commonwealth University, Adult Education

Director: James McMillan
Ph.D., Foundations of Education

Virginia Commonwealth University
Richmond, Virginia
March 2006

Table of Contents

List of Tables

Acknowledgement

I would like to several people who have supported me throughout the last four years. First, I want to thank my husband, Brian, for doing the bulk of the laundry, cooking, and cleaning while I sat at the computer – sometimes typing away, sometimes just staring. I could not have finished this without his love and support. I would also like to thank my parents, my sister and my aunts for their encouragement and constant reminders that I was capable of finishing.

I also could not have completed the process without my dissertation chair, Dr. McMillan, who has supported me every step of the way. The rest of my committee, Drs. Abrams, Rhodes, and Williams, pointed me in the right direction and just wanted me to make my dissertation the best it could be. Joan – if you hadn't given me deadlines (and then reminded me of them), I might still be writing! Finally, I would like to thank the school system that allowed me to use their data.

List of Figures

ABSTRACT

THE RELATIONSHIP BETWEEN THE VIRGINIA STANDARDS OF LEARNING TESTS AND THE NEW PSAT/NMSQT

By Susan P. McKelvey, Ph.D.

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

Virginia Commonwealth University, 2006

Major Director: James McMillan, Ph.D., Department of Foundations

This study examined the relationships between the SOL End-of-Course Reading and Writing tests and the new PSAT/NMSQT Verbal and Writing tests. The PSAT/NMSQT Writing tests were administered for the first time in October 2004. Two linear regression analyses were utilized, with PSAT/NMSQT Verbal and Writing scores, gender, race, and special education as the independent variables, and SOL End-of-Course Reading and Writing scores as the dependent variables. Additionally, two logistic regression analyses were employed with the same variables to predict whether or not a student would pass the SOL End-of-Course Reading and Writing tests. Results indicated that the PSAT/NMSQT Verbal and Writing scores accounted for the bulk of the variance in the SOL Reading and Writing scores. Special education students were predicted to have much lower scores than their non-special education counterparts. Gender and race contributed the least to the regression analyses. With the emphasis on scientifically-based research, this study could be utilized to develop remediation programs for students predicted to fail the SOL tests.

Further research is warranted using additional variables, such as GPA, socio-economic status, and a wider variety of race.

CHAPTER 1

OVERVIEW OF THE STUDY

Background for the Study

Standardized testing is used pervasively in American education, particularly with the recent reinforcement of standards-based accountability. Standards-based testing has recently become the cornerstone for President George W. Bush's educational reform policies. However, this concept is not new. In 1965, Lyndon B. Johnson wanted to involve the federal government to improve educational opportunities for the lower class (National, 2005). President Johnson did this by creating the Elementary and Secondary Education Act (ESEA), which allowed the government to give the most money to public education in the history of the American educational system. In the 1980's, the Reagan administration produced "A Nation at Risk," which showed that most American students would not be able to meet basic standards. This started the standards-based movement. In 2001, President George W. Bush signed the reauthorization of the ESEA, called the *No Child Left Behind Act of 2001.* This Act allows the federal government to be involved even more in public education.

The *No Child Left Behind Act* (2001) expanded state and national education reform efforts. This law purports to narrow and close the gap between white and minority students. The mandates of the *No Child Left Behind Act* demand more accountability from school systems, putting more control at the local level, giving more options to parents, and placing more emphasis on better teaching methods. All states are required to

have standards, as well as a way to measure student learning of those standards. Virginia had standards and assessments before the implementation of NCLB and gave its first required standardized SOL tests in spring 1998, putting Virginia school systems ahead of many others in meeting requirements for NCLB (VDOE, 1999). In addition, the *No Child Left Behind Act* (U.S. Department of Education, 2001) demands scientifically based research to make improvements in the educational system.

Standardized testing is not limited only to standards-based tests. In 1926, the first Scholastic Aptitude Test (SAT), which is a standardized aptitude test, was administered to college-bound high school, mainly male, students (Frontline, 1999). Closely following, in 1959 the American College Test (ACT) was also administered (American, 2004). Both the SAT and ACT are standardized tests purporting to measure general aptitude and are now accepted as part of a student's college admissions application. The reason colleges require one of these tests is because they are both shown to predict student performance in college, particularly during his or her first year (American, 2004; The College Board, 2004b).

Standardized testing has recently been the buzzword in education, with both critics and supporters. The supporters state that tests need to be standardized because standardization allows for a common ground among test-takers, regardless of socio-economic background. Critics state that tests are used improperly, and that they are biased toward those with higher socio-economic status. Many of these questions revolve around validity, which focuses on how the test scores are used, whether they are college

entrance examinations such as the SAT, or statewide tests, such as the Virginia SOL tests.

Overview of the Literature

Use of the SAT and Virginia SOLs

Most colleges establish a minimum score requirement for the SAT, and they use that in conjunction with other items, such as high school grades, admissions essay, and the interview to make final admissions decisions (College Board, 2006b). The PSAT/NMSQT, which students take before the SAT as preparation, is not required for admission into college, nor are students required to take it before taking the SAT. However, scores from the PSAT/NMSQT are a very good indicator of scores on the SAT because the PSAT/NMSQT tests are made up of old questions from the SAT tests. Additionally, the PSAT/NMSQT scores are good indicators of SAT scores because they follow the same format and measure the same constructs (College Board, 2004b). The College Board also states that if students take the PSAT/NMSQT in preparation for the SAT, those students score on average ten points higher on the SAT Reading test and 14 points higher on the SAT Writing tests than they score on the PSAT/NMSQT (College Board, 2006b). In addition, PSAT/NMSQT scores are used to predict performance on Advanced Placement (AP) examinations (Palin, 2001). Students who take the PSAT/NMSQT and SAT receive three scores – a verbal score, a writing score, and a math score (The College Board, 2004a). The highest attainable score on the SAT is 2400, or 800 verbal, 800 writing, and 800 math. The scores on each section of the test range

from 200 to 800. The higher a student scores on the SAT, the better his or her chances of being predicted to succeed during the first year of college. The PSAT/NMSQT is scored similarly, with 80 being the highest attainable score for each test. The highest a student can score on the PSAT/NMSQT verbal, math, and writing tests combined is 240. Even though the scores are different on the PSAT/NMSQT and the SAT, the College Board (2000) suggests adding a zero to the PSAT/NMSQT score for comparison to the SAT. For example, a score of 100 on the PSAT/NMSQT is equivalent to a score of 1000 on the SAT. Camara and Echternacht (2000) found that SAT scores and college freshman GPA are highly correlated, meaning that the higher a student scores on the SAT, the better his or her chances of succeeding during the first year of college. Starting in the 2004-2005 school year, the writing section for both the PSAT/NMSQT and SAT tests was a new test. Previously, the highest attainable scores on the SAT and PSAT/NMSQT were 1600 and 160, respectively, because the tests consisted of only the verbal and math sections (The College Board, 2000). In addition, some changes were made to the types of items on both the math and verbal portions of the test. For example, the verbal tests previously included analogies on the verbal test, which were removed from the tests (The College Board, 2004a).

Since approximately 90 percent of students seeking admission to four-year colleges take the SAT or ACT (ETS, 2000), the College Board must assure that admissions officers are utilizing the scores effectively. Prior to using a test and test scores for any interpretation, test developers must ensure that the tests have evidence supporting the reliability and validity of the scores. Validity evidence answers the question: Does the

test measure what it is supposed to measure? Reliability evidence answers the question: Are the results consistent (Zucker, 2003)? The American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council of Measurement in Education (NCME) (1999) state that validity is, "the most fundamental consideration in developing and evaluating tests. The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations" (p. 9). Furthermore, "When the tests scores are used or interpreted in more than one way, each intended interpretation must be validated" (p. 9). Extensive reliability and validity information has been collected for these two tests. This includes ensuring that the tests are not biased for or against certain groups. Schuppan, Curley, O'Callaghan, and Schmidt (2004) examined the content validity; Cahalan, Mandinach, and Camara (2002) studied the use of SAT scores for special education students; Camara and Echternacht (2000) discussed their findings on the ability of the SAT to predict college grades, particularly when coupled with high school grade point average (GPA); and Frisch-Kowalski (2003) offered a very detailed history of the SAT, including citations of the collection of reliability and validity evidence. The PSAT/NMSQT consists of old SAT questions, which have previously endured the reliability and validation processes (College Board, 2004b).

The same reliability and validity measures must be taken for all other tests, including state-wide mandated tests, such as those necessitated by the No Child Left Behind Act of 2001 (U.S. Department of Education, 2001). In the state of Virginia, a failing score on the Standards of Learning (SOL), end-of-course tests is one reason that a

student may not be able to graduate. Using test scores to make decisions such as whether or not a student moves on to the next grade or graduates from high school is called high stakes testing (Kohn, 2000). While the *No Child Left Behind Act* does not require states to implement high stakes on their standards-based tests, 23 states require passing scores on statewide or end-of-course examinations for graduation (Education Week, 2006). Three additional states will have the same policy in the near future. Finally, while Delaware, Wisconsin, and Missouri do not require passing scores for graduation, they do require passing scores for promotion between grade levels (Education Week, 2006). The 2005-2006 school year was the first year that all students in the state of Virginia took a reading and math SOL test each year in the third through eighth grades. In addition, all students were tested at least once in both reading and math at the high school level. Prior to the 2005-2006 school year, SOL tests were given at third, fifth, and eighth grades, and during high school at the end of each core course. The tests are given in the core subjects of English, math, science and social studies. Students in Virginia are not retained for not passing their SOL tests; however, as stated above, they may not graduate if they do not pass the end-of-course SOL tests during high school because passing scores on the SOL tests allows them to earn verified credits which are required for graduation (Virginia Department of Education, 2003).

Students strive to score well on the PSAT/NMSQT and SAT, as well as their state-mandated standardized tests. While the PSAT/NMSQT is the best indicator of how a student will perform on the SAT (The College Board, 2004b), the SAT is, in turn, is one of the best gauges of a student's first year GPA in college (Camara, 2000; Frisch-

Kowalski, 2003; Frontline, 1999). The PSAT/NMSQT and the SAT both have new writing sections, which were administered for the first time in the 2004-05 school year (The College Board, 2004a). Camara and Echternacht (2000) also state that many validity studies have been done with the SAT that show, along with high school GPA, the SAT is a solid predictor of success in college.

Most students taking the SAT are college-bound students, since the SAT is used at most colleges as an admission criterion. Colleges use the scores from this test because it has been the accepted measure of aptitude since its inception in 1926, as well as the best predictive indicator of college performance, along with high school GPA (Camara, 2000; Frisch-Kowalski, 2003). On the other hand, the Virginia SOL tests are standards-based tests, which were developed to measure whether or not students met the state content standards. All students, regardless of being college-bound, must take these tests, which have been established as an acceptable measurement to fulfill the *No Child Left Behind* requirements (VDOE, 2002).

Statement of the Problem

The SAT is widely used for college admission, and the PSAT/NMSQT is normally used for preparation for the SAT. Both of these tests are used in Virginia. The implementation of the *No Child Left Behind Act* made this a crucial time for using all resources available to help students prepare for the standards-based tests, and is particularly true in Virginia and the other states that attach high stakes to their state-wide tests. An examination of the relationship between the PSAT/NMSQT and the SOL tests

could provide an additional resource for school systems to use, which could help students and schools meet the requirements of NCLB by identifying students in need of remediation. According to Education Weeks' Quality Counts (2006) report, 26 percent of schools did not make adequate yearly progress as required by the *No Child Left Behind Act*. This means that the students in these schools did not pass the required statewide tests, which are based on basic standards in the core curriculum. Additionally, data from the 2004-2005 school year provided evidence that 14 percent of schools in the nation need improvement (Education, 2006). Schools could use PSAT/NMSQT data to develop English and/or writing programs.

An investigation of the literature did not reveal any correlation studies using the PSAT/NMSQT, SAT and Virginia SOL tests. In addition, the 2004-2005 school year was the first opportunity to observe any relationship between the writing portions of the two tests because the PSAT/NMSQT implemented a writing prompt for the first time that year. Though the SAT and PSAT/NMSQT have sufficient data to warrant their predictability of college performance, there has been no research on the correlation of the PSAT/ NMSQT with scores on the SOL tests. At this time, a correlation is particularly important because of the new writing section. This is important because the PSAT/NMSQT can act as a resource for helping schools determine which students need extra help in developing writing skills. Also, the literature did not reveal any research using linear regression with the PSAT/NMSQT (as well as other demographic variables) to predict passing scores on the SOL tests. This could be very important in identifying students who need help in specific areas prior to taking the SOL tests. A similar study by

Jane Warry in 2003, which is discussed in Chapter 2, was conducted in Massachusetts using the PSAT, type of community (urban versus suburban), gender, and race to predict scores on the Massachusetts Comprehensive Assessment System (MCAS).

Because of the recent changes and additions to the PSAT/NMSQT, the literature is somewhat lacking in the area of reliability and validity studies. While the Virginia SOL tests have sufficient evidence of validity and reliability, no links have been made between this and the PSAT/NMSQT. In fact, the only literature found in this review concerning correlations between the SOLs and other standardized tests (the Stanford 9 and the Virginia Literacy Passport tests) was in the SOL Technical Report (2000). However, this is expected because the PSAT/NMSQT and SOL test scores are validated for different constructs, such as general aptitude versus achievement. In addition, one study found a positive correlation between the PSAT/NMSQT and the Massachusetts state standards-based tests. The PSAT/NMSQT has also been positively correlated with the Advanced Placement (AP) tests. In particular, no studies used the PSAT/NMSQT to examine correlations with or predict SOL scores.

At this time, understanding the relationship between the PSAT/NMSQT and the SOL tests could be very valuable to school systems in Virginia because of the reasons stated above. Both of these tests influence the future of the students, directly and indirectly. The PSAT/NMSQT is the National Merit Scholarship Qualifying Test, and the scholarship possibilities as well as the scores can influence whether or not a student will start the college application process, or which types of colleges a student will be able to

consider. Student performance on the end-of-course SOL tests determines whether or not a student will graduate. Consequently, the test acts as a barrier, possibly influencing whether or not a student attends college. When a student does not pass the end-of-course SOL tests, the student does not graduate from high school, and is unable to attend a degree program. Kathleen Porter (2002) gave several reasons why a college degree is so important. First, people who attend college earn more money during their lifetime than those who do not attend college. This leads to the ability to save more money, have a higher quality of life, and the ability to be able to enjoy hobbies and leisure time activities. College graduates also enjoy non-monetary benefits, such as learning more about world events and raising their social status. Porter also discussed studies that have shown a correlation between a college education and good health, which also extends to the college graduate's children. In addition to personal benefits, a society of college graduates enjoys decreased dependence on government support, "increased tax revenues" (p. 2) and a better workforce.

Research Questions

The four research questions for this study are:

1. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict the SOL End-of-Course Reading scores?

2. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict the SOL End-of-Course Writing scores?
3. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict whether or not a student will pass the SOL End-of-Course Reading test?
4. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict whether or not a student will pass the SOL End-of-Course Writing test?

Design and Methods

Using data from a large school division in Virginia, scores from eleventh-grade high school students who took the PSAT/NMSQT in October 2004, as well as their end-of-course SOL tests for English and Writing in May 2005 were analyzed. This school division has just over 58,000 students. In this particular school system, all juniors were required to take the PSAT/NMSQT in the Fall semester, and the English and Writing SOL tests in the Spring semester, resulting in approximately 2,500 sets of scores to analyze. The actual number of students is higher than this; however, some students took the SOL tests earlier than the spring of their junior year, were absent, or were not required to take the tests (certain special education categories). Correlations for all tests were examined in SPSS. A linear regression analysis was conducted to predict SOL scores, based on performance on the PSAT/NMSQT. In addition, a logistic regression

was utilized to predict the probability of passing the SOL tests, also based on PSAT/NMSQT scores. In both regression analyses, residual analysis was performed, and the PSAT/NMSQT Verbal and Writing scores, race, special education, and gender were the predictor variables.

Delimitations

This study has several delimitations. First, only the English and writing portions of both the PSAT/NMSQT and SOL were used. This study could also be conducted with the math tests because students take both the English and math portions of the PSAT/NMSQT at the same time. However, the researcher delimited the study to only English because of the addition of the writing section for this particular year. Second, the researcher will use only four categories of race. Although this school system had eight categories of race, the categories other than White, African-American, and Asian had too few cases to reach any meaningful conclusions in the analyses. The school division also had many categories for special education; therefore, the special education variable consisted of the three categories of "Learning Disabled," "All Other Special Education" and "No Special Education." The results of this study can be generalized to other school systems in Virginia that have similar demographics as the school system in this study.

CHAPTER 2

REVIEW OF THE LITERATURE

This chapter provides a background for the SOL tests and the PSAT/NMSQT, with an historical overview of testing in America to provide context. This section also focuses on standardized and standards-based testing in general. Descriptions of reliability and validity standards are explained, as well as the reliability and validity measures taken with the SOL tests and SAT. Standardized testing, particularly high stakes testing, has been a controversial topic, with both supporters and critics voicing their opinions, particularly related to the use, or validity, of the test scores.

Introduction

Students, parents, teachers, and administrators in American public schools are very familiar with the emphasis on standards-based testing. Standards-based testing has become even more prominent over the last few years, particularly with the standards and accountability movements in our schools, as well as in colleges and universities (Haney, Madaus, & Lyons, 1993; Kohn, 2000; Sacks, 1997). The Elementary and Secondary Education Act (ESEA) of 1965, which started Lyndon B. Johnson's War on Poverty, was implemented to benefit the lower class by providing federal monies for their educational improvement (National, 2005). The Act also prompted the development of the U.S. Department of Education, which Ronald Reagan tried to dismantle during the 1980's. Despite producing a report called "A Nation at Risk," which found that four-fifths of American students were not able to meet basic academic standards, President Reagan also placed major cuts on federal funds for education. Reagan then instructed each state

to develop its own academic standards. By 1990, almost 40 percent of American high school students met the "core curriculum requirements" defined in "A Nation at Risk" (National, 2005), according to a report by the National Center for Education Statistics (NCES). This was the beginning of the standards-based accountability movement. State accountability systems have endured some changes over the years, and the reauthorization of the ESEA became the *No Child Left Behind Act of* 2001, which requires states to comply with its requirements to continue to receive federal funding. Under this law, states must have highly qualified teachers in all classrooms, and they must hold schools accountable for student performance on standards-based tests (National, 2005).

Standards-based, Standardized, and High-Stakes Tests

The standards-based tests that states have developed in their school systems are normally standardized tests. The North Central Regional Educational Laboratory (2004) Web site states that "A standardized test is one that is administered under standardized or controlled conditions that specify where, when, how, and for how long children may respond to questions or prompts" (¶ 1). Standardized tests are interpreted within frameworks, such as norm-referenced and criterion-referenced. Norms are created by testing a randomly selected group, which is nationally representative, and the scores from subsequent administrations are compared to this norm group. A criterion-referenced test is developed based on a set of standards, and a student's score will be at a certain level, such as proficient or advanced. The Virginia SOL tests are considered both norm- and

criterion-referenced tests because they have a state-wide norming group, and the tests are based on the state standards (Zucker, 2003). What exactly is standards-based testing, how often is it used, and for what purposes? This question is important in addressing the research questions because it provides a contextual background for why the SOL tests are currently in place.

Tests that are designed to measure specific standards and/or objectives, which are graded based on the percentage correct, are called criterion-referenced testing. Standards-based tests are usually in the form of a multiple choice test, which allows them to be easily and quickly scored (Kohn, 2000; Zucker, 2003). They are given for many different purposes, such as assessing the need for remediation or class placement. Any subject or construct can be tested in this way – math, science, general aptitude (Zucker, 2003). Classroom teachers administer standards-based tests starting in elementary school through the twelfth grade to assess knowledge in a given subject (Sacks, 1997; Zucker 2003). Other types of tests, such as the SAT and PSAT/NMSQT which are not standards-based, are called norm-referenced tests. Norm-referenced tests have a norming group, and a student is graded based on the scores of the students in the norm group. For example, if a student takes a norm-referenced test and is told that he or she is in the 90th percentile, this means that he or she has scored better than 90% of the students in the norm group. School districts may administer norm-referenced tests to determine giftedness or a deficiency in a certain area. The SAT and PSAT/NMSQT are considered aptitude tests, which means that they measure skills that are developed over a period of time (Frisch-Kowalski, 2003). This is different from the Virginia SOL tests, which measure whether

or not students have learned the standards although some constructs measured on both tests may overlap.

Under the *No Child Left Behind Act of 2001,* each state was required to develop standards, as well as standards-based tests that determine whether or not a child has met those standards. During the Reagan administration, however, many states had developed their accountability systems, and had standards and accompanying tests in place before NCLB (National, 2005). A total of 29 states decided to also use their tests to determine whether or not a student will move on to the next grade, graduate from high school, or both (Education Week, 2006). This is known as high-stakes testing, and the stakes are high for schools, teachers, and administrators as well as for students. For example, ten states will close schools based on student performance on their standards-based tests. On the other hand, 16 states reward schools that produce high or improved scores. Some of those states reward the schools with extra funding, and others give monetary rewards to individual teachers whose students had passing scores (Westchester, 2003). While this is not the sole purpose of the test, it has become a common use of the scores (Kohn, 2000; Westchester, 2003). Another use of tests is for admission into colleges or certain graduate programs (Camara & Echternacht, 2000; Chandler, 1999). Some professions, such as doctors, lawyers, and nurses must pass board exams that are standardized. National tests, such as the National Assessment of Educational Progress (NAEP) and the Third International Mathematics and Science Study (TIMSS) are given to a certain number of public school students for global reporting to see how our students fare against students in other countries (Education Commission, 2006). The NAEP test is given to a sample of

schools in states that have volunteered to participate. The schools that are selected take the fourth and eighth grade assessments in reading and math, and the participating schools receive Title I funding (VDOE, 2006b). The TIMSS tests are given to countries who have volunteered to participate, testing students in math and science. Participation in the TIMSS allows countries to measure their students' achievement, compared to those in other participating countries. The TIMSS is given every four years (TIMSS, 2006).

Criticism of Standardized Testing

Standardized and standards-based testing have been debated topics, particularly since the inception of NCLB. The critics do not necessarily disapprove of the tests themselves; they are critical of the manner in which the scores are used, such as in high-stakes testing. Most of the issues of the uses of high-stakes tests focus on validity of the test scores. The critics do not believe that the scores are being utilized correctly. For example, Hilliard (2000) stated that very few states have collected evidence supporting the validity of their high-stakes test scores. Hilliard elaborated that testing technical manuals claim that the test scores have predictive validity; however, the manuals only report correlations between test scores and achievement. Hilliard also provided anecdotal information concerning children who scored poorly on tests only to obtain success later in life. Hilliard was demonstrating that poor test scores do not necessarily result in later failures in life. In Alfie Kohn's (2000) article, *Burnt at the High Stakes*, Kohn stated that the United States is the only country that does such a great amount of standardized testing, and that students are suffering as a result, because they are forced to simply

memorize material. Kohn and other critics also stated that research that has shown that performance on these tests, particularly those given in the public schools, is directly related to socio-economic status (Gandara & Lopez, 1998; Garcia & Fleming, 1998; Kohn, 2000; Sacks, 1997). These same critics of standards-based testing also believed that certain minority groups will always score behind their white counterparts because of issues of bias.

In addition to high-stakes tests, many educators have criticized the SAT, which is a common admissions test for college-bound students (Freedman, 2003; Gandara & Lopez, 1998; Garcia & Fleming, 1998; Schrag, 2001a). For example, in recent years, Atkinson, the President of the University of California, suggested that the SAT be removed as a requirement for entry into his institution. He believed that the SAT II tests (subject tests) and high school grades are a much better indicator of how a student will do in college (Schrag, 2001a). Approximately 1.3 million students take the SAT each year, and Atkinson believes that the test discriminates against minorities. After looking at the SAT data on University of California applicants, he found that the gap between scores for white versus African-American students is much wider than the gap for the same groups on the SAT II. The President of the College Board stated that changes were made to the SAT because of Atkinson's criticisms (Schrag, 2001b).

Garcia and Fleming (1998) also examined the predictive power, one aspect of validity, of the SAT, comparing white and African-American students. They concluded that standardized tests are not fair to African-Americans. In Garcia and Fleming's study, they found that the SAT predicts college performance for white students, but

underestimates for African-American students. Garcia and Fleming did note the limitation of their small sample sizes for many of the comparisons they made, which could compromise the interpretation of the results. For example, they examined correlations between SAT scores and GPA for college seniors. In this case, they had less than 20 in most categories, and as few as six and seven in others. In this same example, they made the case that SAT scores and senior GPA are not correlated, which are the most common items used to predict college freshman year GPA.

Gandara and Lopez (1998) examined the Latino college-bound population, and they had similar results as Garcia and Fleming. Their study showed that SAT scores did not predict college grades for Latinos. In fact, Latinos have historically not scored as well on standardized tests as their white counterparts. Gandara and Lopez discovered that the Latino students had lower self-esteem resulting directly from doing poorly on the SAT. This study had several problems, including a very low sample size of 48, which can prevent results from being generalizable, especially since most of the participants were from the same school and were not necessarily representative of the Latino population. In addition, the methods were questionable. For example, the freshman college GPA was split into five levels, and the SAT scores were split into two levels (high versus low). The Chi-square analysis showed no relationship between GPA and SAT. A simple correlation between SAT scores and GPA instead of levels of SAT and GPA may have revealed richer results.

Finally, the SAT came under more scrutiny because of a testing accommodation issue. One physically disabled student took the Graduate Management Admission Test

(GMAT) for entry to a business school, and because of his disability, he was given extra time to complete the test. The college to which he was applying was notified of the accommodation by the Educational Testing Service (ETS), the same company that administers the SAT tests, who "flagged" the test. The student sued ETS because he felt as though admissions offices may discriminate against him, and ETS agreed to eliminate flagging on all of its tests, including the SAT (Freedman, 2003). Some educators disagreed with the abolishment of flagging and criticized ETS for its decision. They believed that the scores for students with and without accommodations could no longer be compared. Stephen Sireci (2004), one of the psychometricians involved in the decision to eliminate flagging, states that there is no reason to continue this practice because, through their extensive research, they found that scores earned with accommodations are just as valid.

Support of Standardized Testing

Regardless of the staunch criticism, standardized test scores are used pervasively and receive support. If a standardized test is used properly, the scores will allow the test user to assess students (North, 2004). When test scores have evidence to support validity, reliability, fairness, and lack of bias, they can put students on common ground (Dowling, 2000). In Dowling's (2000) article, *Enemies of Promise: Why America Needs the SAT,* Dowling stated that the SAT was developed to allow those who were capable, yet poor, a chance to be admitted to college. Organ (2001) also supported the use of the SAT because of its predictive powers. In addition, Organ stated that while the SAT should

remain as an admission requirement for colleges, it should not be the only consideration for whether or not a student is admitted. Camara and Echternacht (2000) also found that high school GPA is the best predictor of freshman college grades; however, the addition of using SAT scores significantly adds to the prediction.

In support of high stakes testing in particular, Greene, Winters, and Forster (2003) gathered empirical evidence to compare high-stakes test scores with "low-stakes" test scores. In their article they defined low stakes tests as any test that does not have consequences such as preventing a student from graduating. The examined 5,587 test scores from nine school systems in eight different states. The state of Virginia was used in this study, and scores from the SOL tests were compared with scores on the Stanford 9. The researchers did not want to use college entrance exams such as the SAT or ACT because they test higher-order thinking skills, which is different from standards-based tests. In the nine school systems, they found high correlations between the high and low stakes tests. In Virginia, the correlation between the SOL and Stanford 9 was high, at .77. Additionally, in Florida, which they stated had the most aggressive accountability system, their high and low stakes test scores correlated at .96. Finally, other supports of high-stakes testing believed that the stakes make students and teachers work harder; therefore, making high stakes a factor in closing the achievement gap (Westchester, 2006).

Scientifically-Based Research in Education

To counter the criticism of standardized and high-stakes testing, the Office of Educational Research and Improvement (OERI) awarded a contract to the Educational

Resources Information Center (ERIC) to develop a database called the "What Works Clearinghouse" (U.S. Department of Education, 2002). The clearinghouse was developed in response to one of the key components of the *No Child Left Behind Act of 2001*, which states that the federal education system will be based on research-proven strategies. The What Works Clearinghouse was designed to provide information on research that has the best scientific design, which follows research-based principles. This information is available to the public so that anyone can review the research provided on the Web site and make their own conclusions. Finally, President Bush also signed into law the Education Sciences Reform Act later that year, which replaced the OERI with the Institute of Education Sciences (OERI, 2002).

At the same time, the National Academy of Sciences published a book titled *Scientific Research in Education* (Shavelson & Towne, 2002) in response to the push for scientifically-based research highlighted in the *No Child Left Behind Act of 2001*. The book was written by the Committee on Scientific Principles for Education Research, and was edited by Richard Shavelson and Lisa Towne. The Committee stated that there are six principles for scientifically-based research that should be followed by all disciplines. The principles are:

1. Pose significant questions that can be investigated empirically.
2. Link research to relevant theory.
3. Use methods that permit direct investigation of the question.
4. Provide a coherent and explicit chain of reasoning.

5. Replicate and generalize across studies.
6. Disclose research to encourage professional scrutiny and critique (pages 3-5).

The Committee also believed that to incorporate the six principles into educational practice specifically, "the design must allow direct, empirical investigation of an important question, account for the context in which they study is carried out, align with a conceptual framework, reflect careful and thorough reasoning, and disclose results to encourage debate in the scientific community" (page 6). This information is important to any study being conducted in the educational realm.

Virginia SOLs

Most of the literature found in ERIC and other educational databases deals with standards-based and high-stakes testing in general. The Virginia State Department of Education provides resources concerning the state SOL test, including the parent's guide for the Virginia SOLs (Virginia, 2001). The guide stated that the core standards are in English, math, science, history (including the social sciences), and computer technology. The Virginia Department of Education adopted the SOLs in June of 1995 in the core subjects. The next year, Harcourt Brace Educational Measurement (HBEM) was employed to work with the Virginia Department of Education, as well as other educators, to develop the SOL assessments. In 1998, the Standard Setting Committees established scores required for the different levels of the tests. The levels were "Fail/Does Not Meet the Standard," "Pass/Proficient," and "Pass/Advanced." The state also developed

standards for other subjects, such as fine arts and foreign language, which are not tested, and will not be tested for some time, through the state-mandated testing program (Virginia, 2001).

Tests for the core subjects were administered in the 3rd, 5th, 8th, and high school grades as end-of-course tests. All of the tests were multiple choice, with the exception of the writing portion of the English assessment, which was a composition. In addition to having untimed tests, students had the option of taking the end-of-course tests as many times as needed to pass. If a student wanted to take another test in lieu of the SOL (end-of-course only), he or she could, as long as the substitute test was on the approved list. For example, a student who reaches a certain score on an AP or SAT II subject test does not have to take the SOL for that subject (Virginia, 2001).

Students in Virginia must pass a certain number of SOL tests to graduate (Virginia, 2006). To earn a high school diploma, students must have a certain number of standard credits and verified credits. To receive a standard unit of credit, a student must have at least 140 hours of instructional time in a course, and the student must have met the course objectives – in other words, earned a passing grade. A verified unit of credit is the same as the standard unit of credit, with the additional requirement of earning a passing score on the corresponding SOL test. A student may obtain a standard diploma, which means that they have earned at least six verified credits. An advanced studies diploma requires nine verified credits. Beginning with the class of 2004, students must pass the two English SOL tests (Reading and Writing), as well as four additional tests, which the students may choose. This requirement is the same for students who graduated

in 2005 and 2006. For the 2007 graduating class, students must pass one math SOL, one science SOL, one history/social science SOL, and one student-selected SOL, in addition to the two English SOL tests.

SAT and PSAT/NMSQT

In 1900, the College Entrance Examination Board (CEEB) formed, with the Scholastic Aptitude Test (SAT) first being administered to 8,040 students in 1926. The CEEB was established to form an alliance between college preparatory schools and universities, purporting to create guidelines on college admissions, as well as create a standard curriculum in the secondary schools. Only one year after the first administration of the SAT, the CEEB set the SAT scale at 200 to 800. In 1929, the test was split into two sections, math and verbal. Students received a math score and a verbal score, each ranging from 200 to 800. Students also received a composite score, which consisted of adding the math and verbal scores together (Frisch-Kowalski, 2003).

The 1930's and 1940's saw more changes to the SAT. First, more attention was given to establishing reliability and development of the test. During these two decades, the number of students taking the SAT increased dramatically – so much so that the CEEB had to add additional test dates throughout the year. Until 1937, the SAT was administered only once per year. The group that took the test at the April 1941 administration became the norming group for all subsequent tests, up until 1995; however, scores were re-equated each year. During this time, the test became multiple choice only, which allowed for fast and accurate machine-scoring. The Achievement

Tests, now called the SAT II subject tests, were also developed during this time. In 1948, the Education Testing Service (ETS) was founded, and started taking over the logistics of the test (Frisch-Kowalski, 2003).

The next three decades (1950's, 1960's and 1970's) brought about even more changes. An increasing number of students kept taking the test, and even more test dates were added each year. In 1956, Georgia universities and colleges started to require the SAT as part of the admissions application, which meant that the use of the test was becoming more widespread. Test developers researched the effects of coaching students on taking the SAT, and they developed the Preliminary Scholastic Aptitude Test (PSAT) as preparation for the SAT, which was first administered in 1959. A few years later, the CEEB started to consider the fact that the student population signing up for and taking the SAT was growing more diverse. By 1977, the number of students taking the test increased, which necessitated six administrations per year, during April and June. The content of the SAT began to change in the 1970's, reflecting the multi-cultural society in America. In addition, fewer students took the subject tests because the SAT was so widely used (Frisch-Kowalski, 2003).

The College Board started offering more programs during the 1970's. They even had an agreement with the National Merit Scholarship Corporation, which prompted a name change for the PSAT. The PSAT would be known from then on as the Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test (PSAT/NMSQT). The PSAT/NMSQT is now used to predict scores on the Advanced Placement Program (AP) tests, which were also developed during the 1970's (Frisch-Kowalski, 2003).

During the 1980's, more than 1.5 million students took the SAT – this included students abroad. Students began taking the SAT in the spring of their junior year of high school to help in deciding where to apply for college admission. Because of increased technology and a larger pool of test questions, multiple forms of the SAT started to be given at each administration, which had become very systematic. The ETS made sure that the administration of the test was completely standardized. Students started to be able to request that specific colleges and universities, as well as their high school, would receive their test scores. Because the students had their test scores sent to their high school, the school counselors were able to recommend which colleges they should be applying to, based on their SAT scores (Frisch-Kowalski, 2003).

Throughout the years, the SAT did not change its purpose. The SAT was put into place to aid in the admissions process for colleges, and it was known to predict success in college (Frisch-Kowalski, 2003). To ensure that they would continue to provide American colleges with a useful resource, the College Board brought together a task force in 1986, which included College Board staff, ETS staff, and education professionals from the entire country, to review the SAT and make any necessary changes. The purpose of the task force was to suggest changes to the SAT testing program which would address the social and educational transformations occurring in America. The task force had many ideas and recommendations which were reviewed to make sure they were technically sound. These ideas were also presented to other key stakeholders, such as educators, college admissions personnel, and prominent members of minority

communities. The College Board made changes to the SAT, requiring field testing for reliability and validity measures (Frisch-Kowalski, 2003).

After six years of development and field testing of the updated SAT, in 1994 the College Board introduced the many changes that had taken place. The SAT would no longer be an acronym for Scholastic Aptitude Test. The full test name was changed to SAT I: Reasoning, and the Achievement Tests were changed to SAT II: Subject Tests. The PSAT/NMSQT retained the same name; however, all of the other changes were the same as the SAT I test. The tests had some content changes, and a few minor format changes. A few of the content changes to the verbal portion were that the antonyms section was dropped, the reading passages were lengthened, and the reading questions measured higher-order skills. On the math portion, students could use calculators when they could not in the past, some of the questions were not multiple choice, and more weight was given to data interpretation questions (Frisch-Kowalski, 2003). In 2004, a writing section was added to both the PSAT/NMSQT and the SAT. Scores on the writing section could range from 200 to 800 on the SAT and 20 to 80 on the PSAT/NMSQT, and students then started to receive four scores – math, verbal, writing, and a composite score which was all three test scores added together. A student with a perfect score on all three tests would receive a composite score of 2400 on the SAT and 240 on the PSAT/NMSQT (The College Board, 2006b). The next section addresses the importance of reliability and validity measures, which need to be taken into consideration any time a test is developed or undergoes any changes, such as with the SAT and PSAT/NMSQT.

Validity and Reliability

Validation is the most important process in test development and interpretation (AERA, 1999). In addition, a test or instrument intended to measure a construct, such as achievement or even attitude, must have reliability data (AERA, 1999). This discussion is crucial to this research because this study examines the predictive validity of the PSAT/NMSQT in reference to the SOL tests. The PSAT/NMSQT scores must be valid and reliable to be such a widely used assessment. This review covers validity and reliability relevant to the SOLs and the SAT tests. Since the PSAT/NMSQT is made up of old questions from the SAT tests (College Board, 2004b), most of the reliability and validity studies have been focused on the SAT.

Validity

One of the most fundamental aspects of validity is how the test scores will be interpreted and what the test will measure. The test developer must gather scientific evidence to do this. Validity must confirm the interpretation and the use of the test scores, and it is an ongoing process. If a test developer believes that a test must be changed because the validity evidence does not support interpretation of the test, then validity evidence must be gathered for the new items (AERA, 1999).

Developers need to make decisions about the types of validity evidence that would be the most relevant to their test and how they intend to use the scores. One point of interest is construct under-representation and construct irrelevance. Under-

representation of a construct means that the items do not fully embody what the test is supposed to measure. On the other hand, construct irrelevance is where the test items measure additional constructs. The next steps in test development would be to gather several types of empirical evidence, review the literature, or do an analytical evaluation (AERA, 1999).

The first type of empirical evidence considers the content of the test. Here, the test developer must have a clear basis for the content, such as the themes, the words, format, the types of questions, and the administration of the test. Even if the developer is familiar with the content, he or she should also ask others qualified in the content field to conduct an expert review. This type of validity evidence could be gathered for any kind of standardized test (AERA, 1999).

The next type of evidence assesses how the test scores relate to other variables. One way to gather evidence is to examine convergent and discriminant relationships. Do scores from another instrument that purports to measure the same thing correlate with the test scores from the instrument in development? If so, then the researcher has convergent evidence. On the other hand, one does not want scores to converge with another test that tests something completely different – this would be discriminant evidence, as long as there is no correlation between the two tests. The next evidence, called test-criterion relationships, examines if test scores can predict a behavior or achievement that happens at a later time. This concept can also be used concurrently, where the "predictor and criterion information" (AERA, 1999, p. 14) are gathered around the same time. The last

type of validity relating scores to other variables is generalization, using meta-analysis. In other words, the test user can look at the statistics from previous gathering of validity evidence, where the test was used in comparable situations. This is important because the test user (i.e., school administrators, employers) has the responsibility to make sure validation has occurred for his or her specific uses. Test users are cautioned against using scores for purposes that have not been validated (AERA, 1999).

Again, validity encompasses the use of test scores and the interpretations of those scores. While a test developer and test user must collect evidence, he or she must do so while also following specific standards set forth by the APA (AERA, 1999). The APA describes 24 standards related to test score validation; however, this paper will describe a few examples related to the SAT and SOL tests.

The first standard states that, "A rationale should be presented for each recommended interpretation and use of test scores, together with a comprehensive summary of the evidence and theory bearing on the intended use or interpretation" (p. 17). In other words, the researcher must determine that the intended use is reasonable, and that the test is being used appropriately. The second standard states that, "The test developer should set forth clearly how test scores are intended to be interpreted and used. The population(s) for which a test is appropriate should be clearly delimited, and the construct that the test is intended to assess should be clearly described" (p. 17). When discussing validation evidence, this refers to the scores, not the actual test. In addition,

one must explicitly define how the test will be used and by whom it will be used (AERA, 1999).

The third standard states, "If validity for some common or likely interpretation has not been investigated, or if the interpretation is inconsistent with available evidence, that fact should be made clear and potential users should be cautioned about making unsupported interpretations" (p. 18). The test developer must let the test users know if validity is weak or non-existent with certain interpretations. Finally, the seventh standard states:

> When a validation rests in part on the opinions or decisions of expert judges, observers, or raters, procedures for selecting such experts and for eliciting judgments or ratings should be fully described. The qualifications, and experience, of the judges should be presented. The description of procedures should include any training and instructions provided, should indicate whether participants reached their decisions independently, and should report the level of agreement reached. If participants interacted with one another or exchanged information, the procedures through which they may have influenced one another should be set forth (p. 19).

When using expert review as part of validation, all procedures from beginning to end must be clear and explicit. As with all of the standards, every step in collecting validity evidence should be recorded and revealed to the test users (American, et al., 1999).

Reliability

Reliability of test scores is assessed by examining how constant test scores remain when given to a group of people more than once. Because external, as well as internal factors may affect the way a person responds to test questions, he or she will rarely be one hundred percent consistent from one administration to another. This creates measurement error, which is an indicator of reliability. Measurement error and reliability are inversely related, meaning that the lower the error, the higher the reliability. Another hurdle for gathering evidence of reliability is that standardized tests are no longer completely standardized. Today, the administration of tests can vary as well as the format. For example, students receiving accommodations may take a different format of the test. This may create a larger error of measurement, thus decreasing the reliability of the scores and preventing the test user from making sound assessments (AERA, 1999).

As with validity, all aspects of reliability and errors of measurement should be disclosed. One way to do this is through reporting on the reliability coefficients. The coefficient, a number that represents the consistency of the test scores has three types. The first type of reliability coefficient is called alternate forms, where parallel forms of the same test are administered. The second type, test-retest reliability is found by giving the same test twice to the same group of people. The third type, internal consistency requires only one administration of the test. The test scores are examined to see if relationships exist.

While the reporting of the coefficients is very important, the standard error of measurement, or SEM, is also essential to the user when interpreting scores. All test results have some standard error of measurement. For example, if a student took the same test several days in a row, his scores may vary. Some of the scores may be slightly higher or lower than his actual ability. The difference between his actual score and his theoretical score is called the standard error of measurement. Because the SEM exists for all tests, a student could theoretically score lower on a test than expected. This could result in a "false negative." On the other hand, the student could score slightly higher, which would result in a "false positive." The SEM should be reported and taken into consideration when making decisions based on test scores.

The APA (1999) also has twenty standards for collecting reliability evidence. Again, this research will present a few of these standards that are relevant to the SOL and PSAT/NMSQT tests. The first states that, "For each total score, sub-score, or combination of scores that is to be interpreted, estimates of relevant reliabilities and standard errors of measurement or test information functions should be reported" (p. 31). This is consistent with previous statements about precise reporting of every bit of evidence collected, and the estimates are included in the SOL Technical Report.

The next relevant standard says that, "Each method of quantifying the precision or consistency of scores should be described clearly and expressed in terms of statistics appropriate to the method. The sampling procedures used to select examinees for reliability analyses and descriptive statistics on these samples should be reported" (p. 32).

The SOL Technical Report includes all necessary reliability analyses. Finally, the another standard states, "If local scores are employed to apply general scoring rules and principles specified by the test developer, local reliability data should be gathered and reported by local authorities when adequate size samples are available" (p. 34). For example, since the SOL writing tests are scored by each locality, inter-rater reliability data must be collected to ensure the scoring is consistent.

Validity and Reliability of the SOL Tests

Since 2001, 31 state assessment systems have endured an external alignment review (with two additional in process at this time), including Virginia (Education Week, 2006). The Technical Report for the Virginia SOL tests (2000) was dedicated to clearly explaining validity and reliability evidence of the 1998 test administration. The report included tables representing evidence to support validity and reliability.

The Technical Report included descriptions of how reliability and validity evidence were collected. First, the test developers engaged a large number of content reviewers, considered experts in their respective fields. The reviewers determined if test questions represented the content of the SOLs, and the group made sure the test was at the correct level of difficulty as well as fair. In addition to this group, there was a Bias Review Committee, which was representative of multiple cultures. The bias review ensured that the tests were not biased toward or against any specific groups of students. A Special Forms Review group looked at the tests to make sure that the format of the questions was appropriate for students who are visually disabled. They examined the

Braille, large-print, and audio forms of the tests. All of the members on these committees were trained specifically to identify any possible bias, the difficulty level, and other concerns of item review. They reviewed the multiple choice items in addition to the writing prompts. The groups' reviews validated the content of the test, including themes, words, and format (VDOE, 2000).

The SOL tests were field tested on a sample of Virginia students in the spring of 1997 (VDOE, 2000). In addition to gathering validity and reliability evidence, the developers wanted to help the teachers become familiar with administering the test, while refining administration procedures. Convergent evidence using the Stanford 9 and Virginia Literacy Passport tests was collected and reported in table format in the Technical Report. These two tests were given annually in Virginia, and the national percentile ranks for these two tests were used in the analysis. The correlations ranged from .53 (Algebra I) to .83 (Grade 8 math). These correlations are considered to be moderately high, which provided a good basis for convergent evidence. The Technical Report included when each test was administered, and the number of students or schools, as required by the APA's standards. Dr. S.E. Phillips, a professor from Michigan State University, acted as one of the external reviewers for the 1998 administration of the SOL tests (VDOE, 1999). In reference to the reliability and validity studies conducted with the Stanford 9 and Virginia Literacy Passport tests, he stated, "The substantial correlations with other measures provide supporting validity evidence for the Virginia SOL tests" (p. 9).

The information on reliability was just as complete as the validity evidence in the report. For all of the multiple choice tests, the developers provided the number of items and students, means and standard deviations of the scores, the internal consistency reliability coefficient, the mean raw score, and the SEM for proficient and advanced cut scores (which ranged from .03 to .08). Similar information was provided for the writing prompt scores. There was a distinction made between the two different types of internal consistency coefficients used for the multiple choice items (Kuder-Richardson Formula) and the writing prompts (coefficient alpha), for even further clarification of the collection of reliability evidence. The Kuder-Richardson Formula for the multiple choice items ranged from .81 to .92, and the coefficient alpha for the writing prompts ranged from .80 to .84. These reliability estimates are considered very good (VDOE, 2000).

As previously discussed, the SOL writing tests were locally scored. The test developers trained individuals to score these tests, and the developers provided inter-rater reliability in the Technical Report. Each composition was evaluated by two independent raters. The inter-rater reliability calculations were classified in three different ways. The first classification, exact agreement, included calculations where the two scorers gave the same score. The next is adjacent agreement, which was when the scorers were one point off from each other. Finally, non-adjacent agreement was when two or more points separated the raters' scores. The scale ranged from one to four. The tables in the Technical Report provided the percentages at each level of agreement for all forms of the writing prompt. Perfect agreement ranged from 60.7 percent to 75.1 percent. Adjacent

agreement ranged from 24.8 percent to 39 percent. Non-adjacent agreement occurred less than .6 percent of the time in all cases, with a low of .1 percent (VDOE, 2000).

The test developers employed an independent group, the Technical Advisory Committee (TAC), to review validity and reliability of the SOL tests. The TAC consisted of people from outside the state of Virginia and Harcourt Brace. The TAC reported that validity evidence based on content was ample. They did report, however, that some of the Virginia Standards of Learning were not conducive to multiple-choice questions; therefore, the tests might not cover every aspect of the standards. In addition, the TAC recommended adding more statistical evidence supporting the bias review. Lastly, they applauded the convergent evidence (TAC, 2004).

For reliability, the TAC members stated that the internal consistency coefficients provided supporting evidence of the reliability of the test scores. TAC members reported that coefficients for the writing portion were not reported. The TAC stated that they were "pleased" (p. 7) to see the consistency and accuracy statistics (which included the false negative and positive information), and that these numbers were sufficient to be able to use the scores for placement of students into the proficient and advanced categories. In general, they asked for more detailed information on how the different forms of the test were developed. Although the TAC suggested changes for improving the reliability and validity evidence, they did state that the SOL tests were of high quality (TAC, 2001).

Validity and Reliability of the SAT

On the other hand, the SAT has been administered since 1926, and many changes have been made to the tests over the years (Frisch-Kowalski, 2003; Schuppan, 2004). According to the APA's standards, any time a test changes, new validity and reliability evidence must be collected (American, et al., 1999). The College Board decided in June 2002 that the SAT needed major revisions to maintain its prominence as a college entrance exam (Schuppan, 2004). Previous validity and reliability evidence had been collected and analyzed for past test forms. The Board conducted six field trials with over 250,000 students between 1988 and 1993. At that time, the researchers compared the reliability data that were collected during these years with previously collected data. With respect to validity, they wanted to at least maintain or improve the predictive powers of the SAT (Frisch-Kowalski, 2003).

The College Board called on 15 content experts to revise the math portion of the SAT. The panel of experts represented multiple cultures and backgrounds, consisting of high school and college math teachers. In the future, this group will advise the College Board on changes to the SAT. The verbal portion of the SAT had a similar panel, with content experts representing different cultures and backgrounds. The content review gave the revised test content validity evidence (Schuppan, 2004).

Predictive validity for the SAT I: Reasoning Test was assessed by Cahalan, Mandinach and Camara (2002), using test scores and high school GPA to predict

freshman college GPA from a sample of students taking the test between 1995 and 1998. Because the number and diversity of students taking the test had increased since the previous studies were conducted, validity evidence needed to be gathered again. In that instance, students with learning disabilities who needed the accommodation of extended time were examined. The researchers wanted to predict the college freshman grade-point-average for this type of student, and they cited disproportionate group sizes as a limitation (n of students with accommodations = 241; n of students with no accommodations = 33,771). The researchers provided the mean and standard deviation for the SAT, as well as for high school and freshman GPA. They also provided correlation coefficients. The correlation coefficients allowed the researchers to conclude that freshman GPA for males with learning disabilities was overpredicted, while GPA for females with learning disabilities was accurate.

With the new writing portion of the SAT, reliability estimates needed to be established. Breland, Kubota, Nickerson, Trapani, and Walker (2004) took on this task. They sampled 600 students from each of four ethnic groups (African American, Asian, Hispanic and White) for a total of 2,400 participants. Some of the participants took two SAT II writing tests, and the comparison group took two persuasive writing prompt tests. The research team reported a variety of coefficients, including Pearson Correlation and coefficient alpha. The researchers found that reliability estimates were slightly lower than those found on the writing portion of the SAT II. The Pearson Correlation for the SAT II was .59, compared to .56 on the persuasive prompt. Similarly, the coefficient alpha was .60 on the SAT II and .55 on the persuasive prompt. When reliability estimates are low,

measurement error is increased. High measurement error indicates that a test score may not accurately represent a student's ability. A test user would be cautioned on how they are utilizing these scores.

Very few studies have emerged that correlate the PSAT/NMSQT with the SOL tests. However, the College Board (2004c) examined the alignment of the PSAT/NMSQT, the AP tests, and the SAT with the Virginia SOLs. In fact, they did this for each state's standards. They found that a high percentage of the SOLs were covered on each of the College Board examinations. In the report, they stated that the SAT program can only cover a subset of each state's standards because of the nature of the test, although they do not disclose their method of alignment. A table illustrating the English SOLs that the Verbal test of the PSAT/NMSQT covers, according to the College Board (2004c) is included in the Appendix.

On the other hand, the PSAT/NMSQT was considered to be a good predictor of success on the AP exams. Camara and Millsap (1998) used the PSAT/NMSQT to examine how well the PSAT/NMSQT scores can predict AP success. Their study expanded on an earlier study by Haag in 1983; however, Haag's study did not use course grades or GPA in the equation, and the study had small sample sizes. Camara and Millsap used PSAT/NMSQT scores from 1993 and 1994. They found a strong correlation between the PSAT/NMSQT and AP exam grades in almost all courses. This relationship is stronger than the relationship between AP class grades and AP exam scores, as well as between GPA and number of courses taken and AP exam scores. They also looked at the

time between taking the PSAT/NMSQT and the AP exam. They found no difference in scores from students who waited seven months between tests compared to those who waited 19 months between tests. In addition, they examined gender and racial differences. The correlations between the PSAT/NMSQT scores and AP exam scores were stronger for females than for males. While the ethnic groups in their study were small, the correlations between PSAT/NMSQT and AP exam scores were also strong across all racial and ethnic groups. Finally, a multiple regression analysis revealed that PSAT/NMSQT scores, GPA, and course grades were the best model to predict AP exam scores.

Crouch and Warry Studies

The PSAT/NMSQT was also used in two other studies. In the first study by Crouch (2003), the researcher used the Texas Assessment of Academic Skills (TAAS) and other demographic variables to predict performance on the PSAT/NMSQT. The TAAS was the standards-based test that students take in the state of Texas. Crouch found that the TAAS math scores were instrumental in predicting the PSAT/NMSQT math scores. Both the TAAS math and reading scores were significantly correlated to the PSAT/NMSQT reading scores. Because the students completed the TAAS before the PSAT/NMSQT, the TAAS scores had to be used as predictors for the PSAT/NMSQT.

In the second study, directly related to the proposed research study, Warry (2003) used the Massachusetts Comprehensive Assessment System (MCAS) to predict PSAT/NMSQT scores. Regression analysis was used to predict MCAS performance, with

type of community, gender, race, and PSAT/NMSQT Verbal and Writing scores as the predictor variables. Warry found that a significant relationship existed between the MCAS and the independent variables. Warry (2003) employed stepwise regression and found that the PSAT Verbal score accounted for most of the variance, and concluded that the Race variable be removed from the equation because it did not add to the predictive ability. While Warry kept the Gender variable in the equation, gender contributed the least of the variables remaining. Warry's (2003) methodology is very similar to the proposed research because regression analysis was used to predict scores on the statewide standards-based test for English and Language Arts.

Warry (2003) examined seven school districts in Massachusetts, and provided descriptions for each district which included unemployment rate, minority rate, school dropout rate, free and reduced lunch, and average SAT verbal score. The suburban schools had lower unemployment, a lower minority population, a lower dropout rate, fewer students on free or reduced lunch, and higher scores on the SAT tests, when compared to the urban districts. The subjects were tenth grade students (532 female and 382 male), who took the PSAT/NMSQT in October 1999 and the MCAS in May 2000. Warry could not include all tenth grade students because they were not required to take the PSAT/NMSQT; therefore, the number of participants represented a sample of the population of tenth grade students. Warry also described the MCAS as having multiple choice questions, an open response section where students were asked to write one to two paragraphs or construct a chart, and a writing prompt for a composition. Students may score at passing advanced, proficient, needs improvement, or failing. Students had two

25-minute verbal sections, two 25-minute math sections, and one 30-minute writing skills section. Warry used only scores from the verbal and writing skills sections in the analysis.

Warry (2003) found a strong relationship between the MCAS English and Language Arts test and the five independent variables. Most of the variance (50.2 percent) in the MCAS scores was due to the PSAT/NMSQT Verbal (not including the Writing Skills section) scores. The PSAT/NMSQT Writing, community, and gender accounted for an additional four percent of the variance. When race was excluded from the regression equation, the results did not change, and Warry further found that the gender variable did not add very much to the model's ability to predict MCAS scores. The conclusion was that PSAT/NMSQT Verbal and Writing, and community contributed the most to the variance in the students' MCAS English and Language Arts scores.

Conclusion

With the importance of scientific evidence and standards-based accountability as outlined in NCLB, validity and reliability studies become crucial because states, such as Virginia, are attaching high stakes to standards-based testing programs. If the PSAT/NMSQT scores, gender, race and special education can predict scores on the SOL End-of-Course tests, then resources could be allocated to helping students who are determined to be in need, according to their test scores. Additionally, if a school is in jeopardy of losing accreditation because of poor SOL scores, the proposed study could have a great impact on those schools as well. The PSAT/NMSQT has been used to

determine how well a student will score on the SAT; it has also been used to predict student success in AP courses; finally, the scores are used for scholarship information. The uses stated here are primarily for college-bound students, who have no problems passing the SOL tests. The results of the proposed research could be used to help all students.

Despite the criticism of standardized testing, high-stakes in particular, the SOL test developers did a thorough job of making sure that the tests were used properly. However, some researchers believed that using the SOL tests as barriers to graduation jeopardized the validity of the test use. This concept is known as consequential validity, referring to the social consequences of the test scores (Messick, 1988, as cited in College Board, 2006a). While SOL test developers made every effort to ensure that sufficient evidence exists supporting validity and reliability of the test scores, consequential validity was also a consideration. The same is true with the College Board's PSAT/NMSQT tests.

The proposed study will have implications in the use of PSAT/NMSQT and SOL scores. While the study will focus on one school system, other localities in Virginia can make the determination if they will be able to use these results which would be particularly relevant to other districts that have a similar population. With the addition of the writing section on the PSAT/NMSQT tests, using the results from these tests could inform strengths and weaknesses in writing preparation. Utilizing the results could allow for changes in writing programs, as well as the creation of a list of best practices to be shared with interested parties.

The proposed research will also provide additional evidence supporting validity, which could be added to the Technical Report, if needed. In addition, the breakdown of demographic information, including special education students, will add to the body of literature.

CHAPTER 3

METHODOLOGY

Chapter 3 will detail the sampling design, including the selection of subjects. Because the review of literature necessitated a detailed assessment of the reliability and validity evidence of the PSAT/NMSQT and the SOL, this section will only provide a review. Next is a description of how the data will be analyzed, while specifically addressing the original research questions and hypotheses. Finally, an overview of the limitations and delimitations of the designs will conclude the chapter.

Introduction

This nonexperimental quantitative study examines the relationship between the PSAT/NMSQT Verbal and Writing test scores and the Virginia SOL English end-of-course test scores through linear and logistic regression. The study also provided the data to establish a standard linear regression equation, using the independent variables of gender, race, special education (delineated here as learning disabled, other special education classification, and non-special education), and PSAT/NMSQT scores to predict SOL scores. The linear regression addressed the first two research questions, which were concerned with the extent to which the PSAT Verbal and Writing scores, gender, race and special education predict SOL End-of-Course Reading and Writing scores. A logistic regression equation was also examined, using the same variables to predict the probability of passing the SOL tests. The third and fourth research questions focused on the extent to which the PSAT Verbal and Writing scores, gender, race and special

education predicted whether or not students passed the SOL End-of-Course Reading and Writing tests.

Subjects

A school system in Virginia provided the data to the researcher, which consisted of the PSAT Verbal and Writing test scores, the SOL End-of-Course Reading and Writing test scores, gender, race, and special education category. The school system requires all juniors to take the PSAT/NMSQT each year; therefore, all of the subjects included in the study were eleventh grade high school students, or juniors, who took the PSAT/NMSQT and the SOL English and Writing end-of-course tests providing approximately 2,500 sets of test scores. The students took the PSAT/NMSQT in October 2004, and they took the SOL English and Writing tests in the spring of 2005. As shown in Table 1, the demographic make-up of this school system, County X, is comparable to the state of Virginia. However, the African-American population in County X is significantly higher than in Virginia.

Table 1

Demographic Breakdown of Dataset

		Study (n = 2588)	School System	State of Virginia
Race	White	61% (1580)	57%	62% (NCES, 2004)
	African-American	31% (809)	35%	27% (NCES, 2004)
	Asian	5% (121)	5%	5% (NCES, 2004)
	Other	3% (78)	4%	6% (NCES, 2004)
Gender	Male	48% (1252)	51%	51% (VDOE, 2004)
	Female	52% (1336)	49%	49% (VDOE, 2004)
Special Education	Learning Disabled	6% (151)	6%	14% total* (NCES, 2004)
	All Other Special Education	5% (121)	7%	
	No Disability	89% (2316)	87%	86%

*Note: The NCES report only gives percentage for students with Individualized Education Programs (IEP).

Measures

The PSAT/NMSQT and the SAT are tests that have a very long history as being the accepted measure of academic ability. Collection of evidence of reliability and validity for the scores on these two tests was not a one-time occurrence because the tests, and those who take it, keep changing. Schuppan (2004) discussed the content validity measures for both portions of the tests, and Frisch-Kowalski (2003) cited all of the validity and reliability evidence collected in an historical overview of the PSAT/NMSQT and the SAT. In addition, evidence supporting fairness and lack of bias were collected to examine race for the writing portion (Breland, et al., 2004) and special education (Cahalan, et al., 2002). Although the SOL tests are fairly new, extensive measures were taken to establish evidence of validity and reliability (Virginia, 2000). While the Technical Advisory Committee (2001) made suggestions for strengthening the state examinations, they found that the tests showed ample evidence for reliability and validity. In addition, the Committee's suggested changes were implemented.

Race and special education variables were used in the proposed study to determine if findings were similar to the Breland et al, and Cahalan et al studies. In the current study, the race variable was broken down into four levels – White, African-American, Asian and Other. The "Other" category consisted of Hispanic, Pacific Islander and Unknown. The three races in the "Other" category did not have enough cases to make any meaningful conclusions, which is why they are grouped together. A similar method was taken with the Special Education variable. The variable was broken down into Learning Disabled, Other Special Education, and No Special Education. The school

system has 13 categories of special education, and two of those categories had only one case. In addition, some of the special education categories are not required to take the PSAT/NMSQT or SOL tests. The PSAT/NMSQT scores, race, gender, and special education were used to predict the SOL scores in the two linear regression analyses, and the SOL scores was a continuous variable. The same independent variables was used to predict whether or not a student passed the SOL test; therefore, the outcome was a dichotomous variable. In this study, the researcher used the probability equation for logistic regression, which means that the outcome was a probability, ranging from zero to one. Table 2 shows all of the variables used in this study.

Table 2

Variables Used

Variable Name (SPSS name)	Type	Levels
SOL End-of-Course Reading – Dependent	Continuous (linear regression) Dichotomous (logistic regression)	N/A
SOL End-of-Course Writing – Dependent	Continuous (linear regression) Dichotomous (logistic regression)	N/A
PSAT/NMSQT Verbal	Continuous	N/A
PSAT/NMSQT Writing	Continuous	N/A
Gender	Dichotomous	Male, Female
Ethnicity	Categorical Indicators (logistic regression)	White, African-American, Asian, Other
Special Education	Categorical Indicators (logistic)	LD, All Other Special Ed, None

Procedures

Data were obtained from County X with the approval of the Research Director. The contact person at County X provided the data in a Microsoft Excel file with all student identifiers removed. The PSAT/NMSQT is administered to all eleventh graders in October of each school year. The SOL English and Writing end-of-course tests are taken in the spring of the eleventh grade year. As stated previously, the number of subjects equaled approximately 2,500, which is a sufficient sample for making meaningful conclusions from the data. The possibility exists that not all students classified as a junior in this school system are included in this study. This is because students may have been absent from school on the day the PSAT/NMSQT was administered. In addition, some students may have taken their SOL End-of-Course English test prior to taking the PSAT/NMSQT for some reason. Finally, many students with particular special education classifications are not required to take the PSAT/NMSQT or the SOL tests, such as the MR and ED classifications. Any students who completed the SOL End-of-Course English test prior to taking the PSAT/NMSQT were not included in the analysis because the PSAT/NMSQT was used here as a predictor.

Data Analysis

The data used for this study were existing data, and were obtained from the school system's Programmer Analyst. The Analyst removed the student identification number before sending the Excel file via email. The data were then imported into SPSS by the researcher. Prior to analyzing the data for specific relationships, assumptions were tested

to detect any violations of normality, linearity, and homogeneity of variance. These assumptions were tested using the scatter plots and probability plots of residuals and predicted residuals. In addition, box plots were examined for outliers. The residual analyses for the linear and logistic regression equations are discussed below. Specifically, the researcher was looking for homogeneity of variance, normality of errors, and high leverage or influential data points that could affect the regression line in linear regression.

One assumption was violated. The SOL test scores were not normally distributed; however, standards-based test scores are not expected to have a normal distribution because the goal is to have students pass a basic standards test. The SOL End-of-Course Reading test scores were slightly negatively skewed, which means that more scores were at the upper end of the range than at the lower end. The same was true for the SOL End-of-Course Writing scores, with many students earning a perfect score. The researcher did not eliminate any cases because of the skewness of the SOL scores. The PSAT/NMSQT Verbal and Writing scores were normally distributed.

Linear Regression

To answer the first two research questions, a linear regression analysis was utilized to determine what score a student earned on the SOL End-of-Course Reading and Writing tests, based on the independent variables (PSAT/NMSQT Verbal and Writing, race, gender, and special education). The linear regression model is an equation represented as:

$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$ (Myers, 1990), where $\hat{y}$ is the predicted SOL End-of-Course Reading or Writing score, β_0 is the intercept, $\beta_1 - \beta_5$ are the regression coefficients, $x_1 - x_5$ are the independent variable values, and ε is the model error. The regression coefficients are the expected change in the dependent variable, and they can be positive or negative. The change in the dependent variable is also based on the unit change in the respective x values, with all of the other x values held constant. The linear regression model with multiple independent variables calculates a predicted value for the dependent variable based on the weighted sum of the predictors, or regression coefficients. The results are stated in terms of standardized or unstandardized coefficients. Standardized coefficients allow the researcher to easily compare the effect of the coefficients. Because unstandardized coefficients are in the original units of measure (such as test scores), the researcher was able to see the predicted values in terms of the original unit of measure.

The predictor variables used in this type of analysis can be continuous or dichotomous. Categorical variables are not used because the categories have arbitrary values applied to them. In this study, the categorical variables of race, and special education were converted to dichotomous, indicator (dummy) variables. For the race variable, which had four levels (White, African-American, Asian, and Other), three variables were created – one for African-American, one for Asian, and one for White. The "Other" variable was not recoded because it was used as the referent, which means that all of the other race variables will be compared to the Other variable in the analysis. The same procedure was applied to the special education classification, using No Special

Education as the referent. The dummy variables help to indicate whether or not a particular category should be used in the model. The PSAT/NMSQT test scores were continuous predictor variables, and the SOL test scores were continuous outcome variables.

Prior to the model selection and interpretation, the researcher examined the residuals. Residuals are used to evaluate the difference between the predicted and observed values, as well as to detect whether or not assumptions were violated. If the differences are small, the model is a good fit. The first assumption tested was whether or not the regression residuals were normally distributed. The testing of the normal distribution was done through a Normal P-Plot. Any deviation from the line indicates deviation from the normal distribution. The P-Plot showed no violation of this assumption. For the next assumption, homogeneity of variance, a scatterplot of the standardized residuals against the standardized predicted values was examined. The plot showed a random scatter of the data points around zero, and no identifiable trends in the data points. For example, the data points showed no pattern, such as funneling, which indicates that the error variance and measured response systematically become larger. The scatterplot and Normal P-Plot are shown in Chapter 4.

Logistic Regression

To answer the third and fourth research questions, logistic regression analysis was used. Logistic regression is similar to linear regression in reference to making predictions. However, logistic regression predicts group membership, or determines the

likelihood of being a case or non-case. In this study, the researcher examined the logistic model to predict the probability of students passing their SOL End-of-Course Reading and Writing tests. Logistic regression is different from linear regression because it does not assume that the residuals of the variables are normally distributed, nor do they have a linear relationship. Logistic regression follows an S-shaped curve, as opposed to a straight line in linear regression. At low levels of the independent variables, the probability is near zero, and at higher levels, the probability continues to increase, but it never reaches one. Myers (1990) states that "the probability of a success is postulated to be a function of a set of regressor variables" (page 317). He also states that this type of regression is commonly used in the social sciences. The model used in this study is represented by

$$\hat{p} = e^{(A + B1X1+B2X2+B3X3)}/(1 + e^{(A+B1X1+B2X2+B3X3)}),$$

where $\hat{p}$ is the predicted probability, A is the constant, B1-B3 are the regression coefficients, and X1-X3 are the values for the predictors. The regression coefficients are calculated through maximum likelihood estimation which makes inferences based on the probabilities of a data set.

In the logistic regression analysis, the outcome variable is the probability that a student will pass the SOL tests. Probability values lie between 0 and 1. Logistic regression uses logarithms and exponents in the calculations, with the focus on natural logarithms. The natural logarithm of a number is the power to which one raises "e" to get that number. The "e" term was used in the probability equations in this analysis. The

same predictor variables will be used as in the linear regression models, and the categorical variables will have the same indicator values assigned.

Research Questions 1 and 2

Once the data set had been examined, the researcher continued with analysis of the data to answer the first two research questions:

1. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict the SOL End-of-Course Reading scores?
2. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict the SOL End-of-Course Writing scores?

The first null hypothesis that none of the independent variables are related to the dependent variables was addressed through the Global *F*-test. The *F*- test shows whether or not relationships are significant. The result was based on an alpha level of .05, and the test was significant at the .05 level. After determining that the *F*-test statistic was significant, the individual predictors were examined. In this standard linear regression model, the null hypothesis is that a particular predictor is not significant in the presence of all other predictors.

The *t*-statistic was then calculated. In the case of the *t*-test, the margin of error is determined, which is called the critical value. The difference between the sample value and the null value is then calculated and divided by the standard error. If the *t* statistic is

higher than the critical value, the researcher determines that the value of the coefficient is significantly different from zero. The conclusion is that the predictor is needed, in the presence of all other predictors. The next step is to conduct a hierarchical multiple regression, which is done through the partial *F*-test. The partial *F* is used to compare models. For example, if a predictor is added to the model, the partial *F* is examined. If the p-value is less than .05 (alpha level set at .05), then the researcher determines that the difference between the two models is significant.

After calculating the Global *F*-test and determining that the result was significant, the *R*-square and Adjusted *R*-square were examined to determine the model's goodness-of-fit. The goodness-of-fit statistics aid the researcher in determining how well the regression equation fits the data. These values are between zero and one (the closer to one, the better). For example, if the *R*-square equals zero, the researcher knows that the independent variables do not help at all in predicting the outcome. The individual predictors, or regression coefficients show the rate of change in the dependent variable (SOL score) as the independent variables increase or decrease. The last step was to see if the predicted dependent values matched the observed values. The closer these two values, the better the model.

Research Questions 3 and 4

After determining the best model using standard and hierarchical linear regression, the researcher conducted a logistic regression to answer the third and fourth research questions:

3. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict whether or not a student will pass the SOL End-of-Course Reading test?
4. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict whether or not a student will pass the SOL End-of-Course Writing test?

In Virginia, students receive a score between zero and 600 on the SOL tests. If the student obtains a score of 400 or above, he or she has passed the test. If the student receives a score of 500 or above, he or she has earned a passing advanced score. The passing advanced classification will not be considered in this study. A logistic regression equation yields the probability that a student will or will not pass, based on the same predictors as in the linear regression. The Wald test was used to test and interpret the coefficients.

Logistic regression is different from linear regression in that the outcome variable is dichotomous. Logistic regression uses the Model Chi-square value, which is a non-parametric test. The Model Chi-square is a value based on degrees of freedom, and it tests the null (nil) hypothesis that all of the regression coefficients are zero, or in other words, have no relationship with the dependent variable (excluding the constant). If the significance level is less than .05, the null hypothesis is rejected, and the researcher concludes that the regression coefficients do not equal zero.

After finding that the Chi-square yielded significant results, the individual predictors were examined. The null hypothesis is that an independent variable is not

significant in the presence of all the other independent variables. The statistic for this is called the Wald test. The Wald test has a z-distribution, which is the parameter estimate divided by the standard error, and the rule of thumb is that if the value of the Wald test is greater than 1.96, then it is significant. The block Chi-square values were also examined which is used to compare models.

The logistic regression model needed to be tested, just as the linear regression model was tested. The goodness-of-fit for the model was done through examining the Nagelkerke F-square whose values range from zero to one. The percent correct predictions were examined to see how many times the model made the correct prediction on the observed variables.

CHAPTER 4

RESULTS

Chapter 4 describes the data analysis in detail, beginning with the descriptive statistics section, which includes frequencies, means, and standard deviations, to correlations and the linear and logistic regression models. The first section is comprised of a description of the preliminary analysis of the data. The next section discusses the linear regression analysis followed by the logistic regression analysis. Chapter 4 closes with a discussion of the limitations of the study.

Data Analysis and Results

This study was designed to answer the following four questions:

1. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict the SOL End-of-Course Reading scores?
2. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict the SOL End-of-Course Writing scores?
3. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict whether or not a student will pass the SOL End-of-Course Reading test?

4. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict whether or not a student will pass the SOL End-of-Course Writing test?

The null hypotheses that were designed to help answer these questions are as follows:

1. There is no significant relationship between the Virginia SOL End-of-Course English/Language Arts scores and the independent variables (PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification).
2. The coefficient values equal zero.

The first step in analyzing the data was to check for violation of assumptions. In linear regression, the researcher must check for missing data, outliers and normality. (These assumptions were not checked for the gender, race, or special education variables because they are not continuous variables, and therefore, are not expected to have outliers or a normal distribution.) Logistic regression does not require the residuals to have a normal distribution; however, outliers need to be carefully evaluated.

As stated previously, some of the data were missing; however, they appeared to be missing at random. The researcher chose to exclude cases listwise, which means that a case would not be considered in the analysis if any data were missing from any of the variables used in the analysis. No transformations were performed on the missing data, although a linear interpolation could be useful. Boxplots and stem and leaf plots were used to detect outliers and extreme values. Outliers can be problematic with any set of data because they raise the risk of Type I or Type II errors, which could result in

problems with generalizations. The SOL End-of-Course Reading test had a few outliers; however, no extreme values were detected that would highly influence the regression line.

The SOL End-of-Course Writing test had a few extreme cases, which might influence the regression analysis. The extreme cases were those where students earned a perfect score of 600 on the Writing test. Extreme cases can bias the regression results by influencing the regression line in a certain direction, and will force the line to pass through the extreme case. The PSAT/NMSQT test scores were normally distributed. The normal probability plots of the residuals were also examined, and the distributions were normal for all of the test scores. The skewness and kurtosis values for the continuous variables were also low. Table 3 shows the skewness and kurtosis data for the four test score variables.

Table 3

Skewness and Kurtosis of SOL and PSAT/NMSQT Test Scores

Variable	Skewness	Kurtosis
SOL End-of-Course Reading	-.010	.021
SOL End-of-Course Writing	.104	-.429
PSAT/NMSQT Verbal	.015	-.400
PSAT/NMSQT Writing	.554	-.408

Skewness is a measure of the symmetry of the data, while kurtosis is a measure of the peakedness or flatness of the data. Extreme values are considered to be greater than three or less than negative three (Data and Statistical Services, 2006).

Descriptive Statistics

The following table gives a description of the population of eleventh graders represented in this study in School District X.

Table 4

Frequencies of Race, Gender, and Special Education

		Frequency – Juniors in study (n = 2588)	Frequency – Juniors total (n = 3128)	Percent in Study (% total)
Gender	Male	1252	1535	48 (49)
	Female	1336	1593	52 (51)
Race	White	1580	1766	61 (57)
	African-American	809	1087	31 (35)
	Asian	121	147	5 (5)
	Other	78	128	3 (4)
Special	LD	151	199	6 (6)
Education	Other	121	233	5 (7)
Classification	No SPED	2316	2696	89 (87)

The levels for each independent variable have a sufficient number of subjects for analysis.

Table 5 shows the descriptive statistics for the SOL End-of-Course Reading and Writing tests, as well as the PSAT/NMSQT Verbal and Writing tests.

Table 5

Descriptive Statistics: SOL Reading and Writing, PSAT/NMSQT Verbal and Writing

Test	Mean	Standard Deviation
SOL End-of-Course Reading	482.55	54.75
SOL End-of-Course Writing	482.04	60.24
PSAT/NMSQT Verbal	44.61	11.50
PSAT/NMSQT Writing	48.78	11.56

Both of the SOL tests scores have a possible range of 0 to 600. The means reported in Table 5 are the arithmetic means (total of scores divided by number of cases). The researcher concluded that students in this sample scored similarly on the SOL End-of-Course Reading and Writing tests. The PSAT/NMSQT test scores could range from 20 to 80. Students in the sample scored higher on the PSAT Writing test. The average scores for the state of Virginia during the same administration of the PSAT/NMSQT were 47.2 on Verbal and 49.7 on Writing, which are a little lower than the scores from the sample of student in this research. Table 6 shows the percentages of students that passed and failed the SOL End-of-Course Reading and Writing tests.

Statistical Analysis for Research Question 1

Linear regression analysis was employed to answer the first research question: To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict the SOL End-of-Course Reading scores?

The statistical analyses were completed in SPSS (Statistical Package for the Social Sciences) 13.0. In SPSS, the researcher entered only the PSAT/NMSQT variables in the first block of the regression analysis. The other demographic variables were grouped together and added to the next blocks. In the second block, the Male, Race, and Special Education variables were added. The tables and brief explanations below show the analysis performed to address the null hypothesis that no significant relationship exists between the SOL Reading test scores and the independent variables. Table 6 includes the model summary information from the models for the stepwise regression performed, with reading SOL as the dependent variable.

Table 6

Model Summary Statistics with SOL Reading as the Dependent Variable

Model	*R*	R^2	Adj R^2	R^2 Change	Standard Error of the Estimate	*df* change	*F* change
1	.728	.530	.529	.530*	36.98	2	1333.06*
2	.736	.541	.540	.011*	36.57	6	9.822*

P-Value > .001

In the two models, the adjusted R-square, indicating the goodness-of-fit, increases one percent from Model 1 at 53% to Model 2 at 54%. Model 1 includes the

PSAT/NMSQT Verbal and Writing scores as the only predictors. In Model 2, the adjusted R-square shows that the group of independent variables account for 54% of the variance in the SOL Reading scores. Model 2 includes both of the PSAT scores, male, Asian, African-American, White, LD, and Special Ed Other variables. The Global *F*-test addresses the null hypothesis that there is no significant relationship between the group of independent variables and the dependent variable. The *F*-test is significant at the .05 level; therefore, this null hypothesis is rejected.

Model 2 includes the predictors that account for the most variance in the dependent variable. The R^2 Change statistic also indicates that the change in the *F*-test was significant from Model 1 to Model 2; therefore, the researcher chose this as the best model even though the demographics do not appear to add much to the equation. The following table shows the coefficients, standard errors, and the *t*-values for each predictor in this model.

Table 7

Coefficients for Model 2

Model	*B*	Standard Error	*t*
(Constant)	323.738	5.995	53.997*
PSAT Verbal	2.119	.119	17.808*
PSAT Writing	1.242	.115	10.840*
Male	-1.167	1.529	-.763
African-American	.230	4.965	.046
Asian	7.020	5.781	1.214
White	8.547	4.915	1.739
LD	-18.095	3.410	-5.307*
SPED Other	-16.311	3.702	-4.405*

**P-Value* < .001.

The coefficients in this table indicate that for every unit increase in PSAT Verbal score, the SOL Reading score will increase by 2.119. This same logic applies to the PSAT Writing scores. The standard error addresses the null hypothesis that the coefficient equals zero. The *t*-test shows that there is a significant difference between the coefficient and zero, and this null hypothesis is rejected. The following is the regression equation to predict the SOL Reading scores:

SOL Reading Predicted = 323.738 + 2.119(PSATVerbal) + 1.242(PSATWriting) – 1.167(Male) + .230(African-American) + 7.020(Asian) + 8.547(White) – 18.095(LD) – 16.311(SPED Other) + Error.

The predictor variables that were converted to dummy variables are included in the equation. The dummy variables are interpreted in comparison to the referent variable. For example, the Special Education referent variable was No Special Education. The LD classification regression coefficient was -18.095, which means that a student classified as LD is predicted to score 18.095 units less than a student who is not classified as LD. The special education classifications have large regression coefficients and are significantly different from zero. On the other hand, the other indicator variables are not significant in the presence of the other predictors. Additionally, while 54% of the variance in SOL Reading score is explained by the model, 46% of the variance is not explained. A goodness-of-fit statistic of 54% is not very strong, and the researcher would caution someone against using the equation. However, the PSAT Verbal and Math scores alone are good indicators of how a student will score on the SOL End-of-Course Reading test. The unexplained variance could be due to influential variables that were not included in the analysis, such as GPA, which will be discussed in the limitations section.

The following figure is the residual plot for the SOL End-of-Course Reading variable.

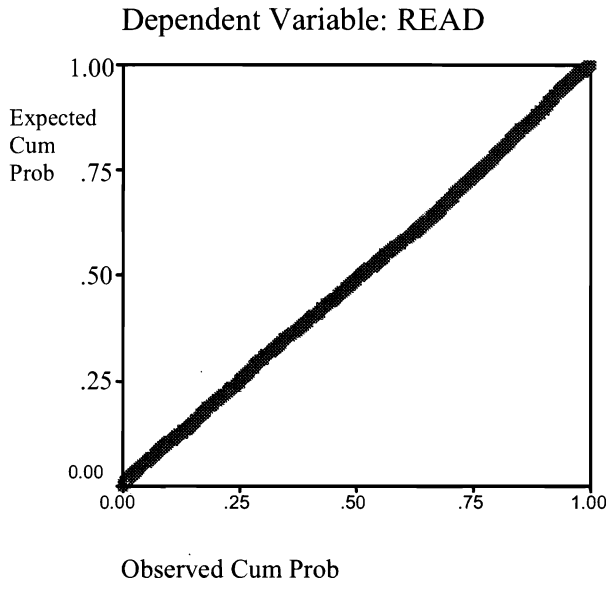

Figure 1. Normal P-P plot of regression standardized residual for SOL End-of-Course Reading test scores.

The Normal P-P plot tests the normal distribution of the residuals. As shown in Figure 1, the residuals are normally distributed, which shows that the assumption of normal distribution of residuals has not been violated. The collinearity diagnostics were also examined. Collinearity can be a problem if two or more predictor variables are highly correlated and could be completely predicted by each other. Collinearity among predictor variables can lead to inaccuracies with the regression analysis. The variance inflation factors (VIF) can help to diagnose this problem. In this case, the VIFs were lower than ten, which indicates no problem with collinearity.

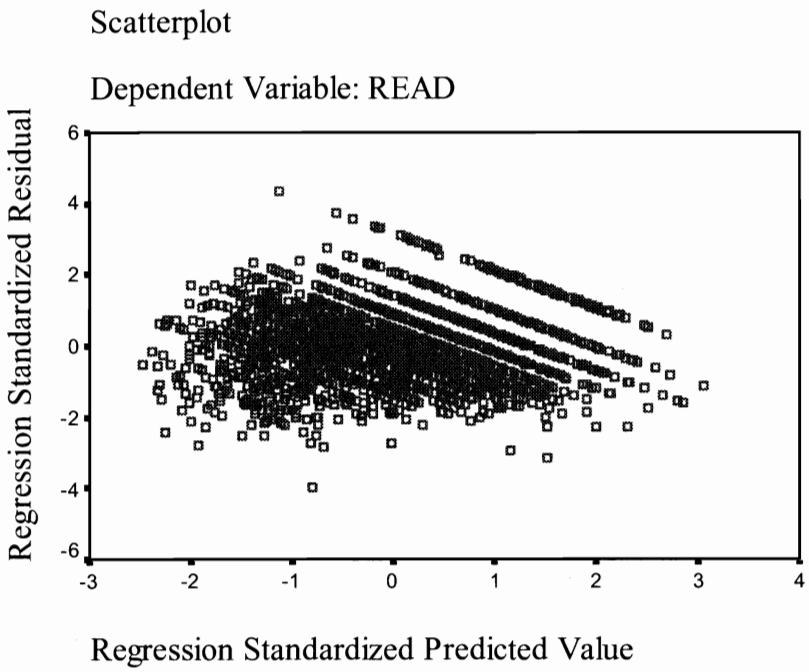

Figure 2. SOL Reading Scatterplot of Regression Standardized Residuals and Standardized Predicted Values.

The scatterplot in Figure 2 shows the standardized residuals and standardized predicted values. The plot shows a fairly random scatter of the data points around zero, with no identifiable trends in the data points, such as curvature or funneling. The data do not violate the homogeneity of variance assumption.

Statistical Analysis for Research Question 2

Linear regression was used again to answer the second research question: To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict the SOL End-of-Course Writing scores? This research question addresses the prediction of the SOL End-of-Course Writing test. The following tables and brief discussions show the relevant statistics for determining the best model to answer this question.

Table 8

Model Summary Statistics with SOL Writing as the Dependent Variable

Model	*R*	R^2	Adj R^2	R^2 Change	Standard Error of the Estimate	*df* change	*F* Change
1	.740	.548	.548	.548	39.048	2	1412.995*
2	.748	.560	.559	.012	38.574	4	15.403*
3	.761	.578	.577	.019	37.769	2	51.075*

P-Value < .001

A similar Model block regression analysis was performed for the second research question. In this analysis, the PSAT Verbal and Writing scores were entered in the first block. In the second block, the Male and Race variables were entered; in the third block, the Special Education variables were entered. The Global *F*-test was significant for all blocks, the *F* Change statistic was significant between all three blocks; and the null

hypothesis that the group of predictors are not significantly related to the dependent variable is rejected.

In the third block of the model, the adjusted R-square was 57%, which means that the model explained 57% of the variance in the dependent variable, SOL Writing test score. As in the first linear regression with SOL Reading test score as the dependent variable, the Adjusted R-square goodness-of-fit statistic explains the majority of the variance; however, 43% of the variance is left unexplained. Again, this may be due to the absence of other variables that may affect the variance, such as GPA. The researcher chose Model 3 as the best regression model because it explains more of the variance than the other two blocks, and the next table shows the coefficients for the equation and their respective *t*-values.

Table 9

Coefficients for Model 3

Model	*B*	Standard Error	*t*
(Constant)	316.840	6.002	52.79*
Verbal	1.371	.125	11.00*
Writing	2.223	.121	18.38*
Male	-8.637	1.606	-5.38*
African-American	-6.002	4.929	-1.22
Asian	7.988	5.820	1.37
White	8.350	4.906	1.70
LD	-28.870	3.668	-7.87*
SPED Other	-23.115	3.853	-6.00*

P-Value < .001

The equation for this model is:

SOL Writing Predicted = 316.840 + 1.371(PSATVerbal) + 2.223(PSATWriting) – 8.637(Male) – 6.002(African-American) + 7.988(Asian) + 8.350(White) – 28.870(LD) – 23.115(SPED Other) + Error.

The coefficients show that for every single increase in PSAT Verbal, there is a 1.371 increase in the SOL Writing score. On the other hand, if the student is a male, his SOL Writing score would decrease by 8.637. In addition, students with classified as Learning Disabled are predicted to have their scores decrease by 28.870 on the SOL Writing test, in comparison to students with no special education classification. The goodness-of-fit

statistic does not indicate that the model is an excellent fit; therefore, the researcher would caution someone against using the regression equation; however, the PSAT Verbal and Writing scores are good indicators of how students will score on the SOL Writing test.

Figure 3 is the residual plot for the SOL End-of-Course Writing variable.

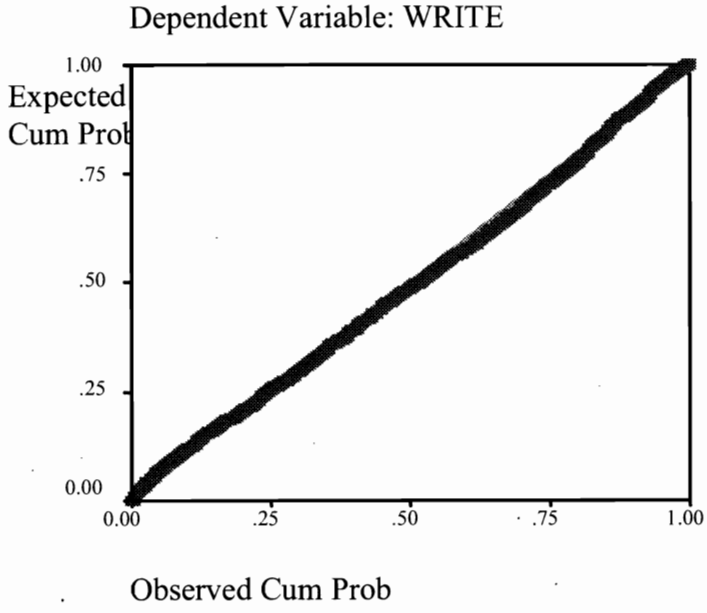

Figure 3. Normal P-P plot of regression standardized residual for SOL End-of-Course Writing test scores.

Again, as shown in Figure 3, the residuals are normally distributed, which shows that the assumption of normal distribution of residuals has not been violated. Collinearity diagnostics were also examined for this research question. There were no VIFs larger than 10.

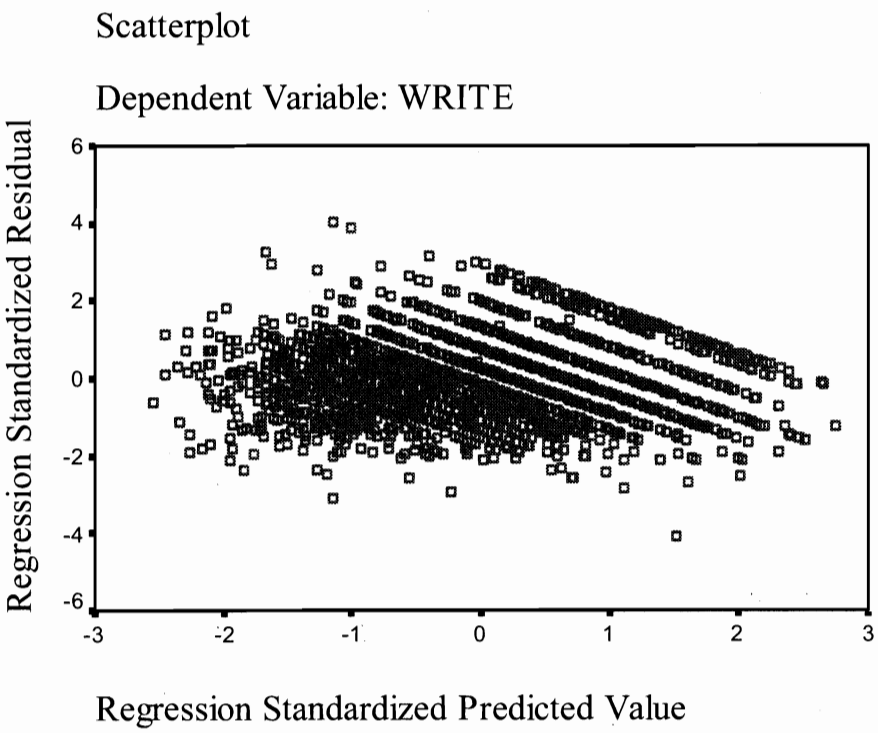

Figure 4. SOL Writing Scatterplot of Regression Standardized Residuals and Standardized Predicted Values.

Statistical Analysis for Research Question 3

The third research question was answered through logistic regression analysis. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict whether or not a student will pass the SOL End-of-Course Reading test? The logistic regression predicted the log-odds of whether or not a student will pass the SOL Reading test. The SOL End-of-Course Reading test scores were transformed into a dichotomous variable (0 = not pass; 1 = pass) for the analysis.

The student demographic predictors were first examined for a significant relationship with the dichotomous dependent variable. All of these demographic variables

are either categorical or dichotomous; therefore, the nonparametric Chi-Square test was utilized for determining which predictors were significantly related to the dependent variable. If the predictors were not significantly related, then they were dropped from the analysis. The following table shows the Chi-square values and significance for each of the student demographic values.

Table 10

Cross-tabulations and Chi-Square Values of Student Demographic Variables - Reading

Variable		Pass Reading	Fail Reading	Chi-Square	*p-value*
Gender	Male	93% (1160)	7% (91)	2.413	.120
	Female	94% (1258)	6% (77)		
Race	White	97% (1526)	3% (54)	74.613	.000
	African-American	88% (706)	12% (101)		
	Asian	95% (115)	5% (6)		
	Other	91% (71)	9% (7)		
Special Ed	LD	74% (111)	26% (39)	202.461	.000
	Special Ed Other	75% (91)	25% (30)		
	No SPED	96% (2216)	4% (99)		

Excluding gender, all of the variables indicated a significant relationship. The non-significant Chi-square value showed that there is no difference between males and females in passing the SOL End-of-Course Reading test. Therefore, this variable was not used in the logistic regression analysis.

The same dummy variables used in the linear regression analyses will be used in the logistic regression. In addition, the Nagelkerke R-Square statistic will be examined as part of the SPSS output; however, researchers are cautioned against using this to explain

the variance in the dependent variable because there is no R-Square equivalent to linear regression.

In the model, the predictor variables in this analysis were entered in four different blocks. In the first block, the PSAT/NMSQT Verbal and Writing scores were entered. The second block included the race variables, and the third block included the special education classification variables. The gender variable (Male) was not used in the analysis because cross-tabulations and Chi-square tests indicated that it was not significantly related to whether or not a student passes the SOL End-of-Course Reading test. In this model, the Chi-Square statistic was run for each block to address the question: Is the model significant? The following table shows the model summary statistics.

Table 11

Model 1 Statistics for SOL End-of-Course Reading – Logistic Regression

	Chi-square change	*df* change	*P*	Nagelkerke R-square	Percent Predicted
Block 1	359.914	2	.000	.395	94%
Block 2	2.957	3	.398	.398	94%
Block 3	32.779	2	.000	.431	94.5%

In addition to the Chi-Square Omnibus Tests of Model Coefficients, another type of Chi-Square was examined for all Blocks. The Hosmer and Lemeshow Test is a goodness-of-

fit test. The goodness-of-fit test tests the null hypothesis that the predicted and observed values are the same. Because none were significant, the researcher did not reject the null hypothesis, and concluded that the predicted and observed values were similar, indicating that the model was a good fit. Table 12 shows the significance of the predictors in Model One, Block Three.

Table 12

Predictor Statistics for Model One Block Three - Reading

Variable	B	Wald Test	*p*	Exp(B)
Constant	-7.458	50.448	.000	.001
PSAT Verbal	.113	47.809	.000	1.119
PSAT Writing	.138	29.572	.000	1.148
African-American	.534	1.120	.290	1.705
Asian	.764	1.346	.246	2.148
White	1.072	4.236	.040	2.920
LD	-1.175	17.605	.000	.309
Special Ed Other	-1.520	23.803	.000	.219

Most of the predictor variables were significant at the .05 level. The race variables of African-American and Asian, however, were not significant in the presence of the other variables. The LD and Special Education Other variables have a negative sign indicating that the probability of a student passing the SOL End-of-Course Reading test decreases

when having a special education classification. This finding is similar to that found in the linear regression analysis. The equation is for this model is:

$$p\hat{} = e^{(-7.878 + .110(\text{PSAT Verbal}) + .148(\text{PSAT Writing}) + ..438(\text{Male}) + .473(\text{African-American}) + .651(\text{Asian}) + .974(\text{White}) - 1.247(\text{LD}) - 1.577(\text{SPED Other}))} / 1 + e^{-7.878 + .110(\text{PSAT Verbal}) + .148(\text{PSAT Writing}) + ..438(\text{Male}) + .473(\text{African-American}) + .651(\text{Asian}) + .974(\text{White}) - 1.247(\text{LD}) - 1.577(\text{SPED Other}))}$$

The coefficients (B) are in log-odds units, and the coefficients give the researcher an indication of the relationship between the predictor variables and the dependent variable. In logistic regression, the dependent variable is on the logit scale, and the log-odds coefficients indicate how much the predict log-odds of the dependent variable will increase or decrease, depending on the sign of the coefficient. The log-odds are difficult to interpret; therefore, they are often changed into odds ratios. The odds ratios are shown in the Exp(B) column. The Exp(B) values tell how much the odds of being a case are multiplied when the independent variable increases by one unit. In the logistic regression equation, the log-odds coefficients were exponentiated, and the predicted variable was in terms of a probability of passing the SOL End-of-Course Reading test.

The researcher also examined the classification table for the model, which showed how many cases were correctly predicted. In block three, 94.5% of the cases were correctly predicted, indicating a good fit. However, 5.5% of cases were incorrectly predicted. The researcher then examined the studentized residuals over two, which was important in detecting outliers. All of the cases with studentized residuals over two were

misclassified as passing the SOL Reading test, when they should have been classified as failing the SOL Reading test. Most were male, and approximately one-half were White and one-half were African-American. A very small amount of the cases were special education students.

Statistical Analysis for Research Question 4

Logistic regression analysis was also used to answer research question four: To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict whether or not a student will pass the SOL End-of-Course Writing test? The same analysis used for question three was used here to predict whether or not a student will pass the SOL End-of-Course Writing test. The SOL End-of-Course Writing test scores were also transformed into a dichotomous variable (0 = not pass; 1 = pass) for the analysis.

Before conducting the logistic regression analysis, the student demographic predictors were examined for a significant relationship with the dependent variable All of these demographic variables are either categorical or dichotomous; therefore, the nonparametric Chi-Square test was utilized for determining which predictors were significantly related to the dependent variable. The following table shows the Chi-square values and significance for each of the student demographic values.

Table 13

Cross-tabulations and Chi-Square Values of Student Demographic Variables - Writing

Variable		Pass Reading	Fail Reading	Chi-Square	*p-value*
Gender	Male	91% (1184)	9% (120)	13.912	.000
	Female	94.5% (1314)	5.5% (76)		
Race	White	97% (1547)	3% (46)		
	African-American	85% (752)	15% (129)	124.242	.000
	Asian	92% (122)	8% (10)		
	Other	88% (77)	12% (11)		
Special Education	LD	72% (105)	28% (41)		
	Special Education Other	77% (99)	23% (29)	190.896	.000
	No SPED	95% (2294)	5% (126)		

In this analysis, all of the variables indicated a significant relationship. The Chi-square value showed that there was a difference between males and females in passing the SOL End-of-Course Writing test (Females are more likely to pass), which was different from the first logistic regression analysis. All of the other variables were significant as well; therefore, all variables were used in the analysis.

In the model, the predictor variables were entered in three different blocks. In the first block, the PSAT/NMSQT Verbal and Writing scores were entered. The second block included the special education classification variables, and the third block included gender (male) and race variables. The researcher decided to enter the special education classification variables before the gender and race variables because the previous analyses indicated that the gender and race variables were not significant in the presence of the other predictors. The following table shows the Model Summary statistics.

Table 14

Model 1 Statistics for SOL End-of-Course Writing

	Chi-square change	*df* change	*P*	Nagelkerke R-square	Percent Predicted
Block 1	330.900	2	.000	.389	94.8%
Block 2	39.748	2	.000	.432	95.2%
Block 3	5.858	4	.210	.439	95%

The percent correctly predicted did not change from Block Two to Block Three, and the Chi-square change was not significant from Block two to Block Three. Again, the Hosmer-Lemshow goodness-of-fit tests for all Blocks were not significant, indicating a good model fit. The researcher chose Block Two as the best model because of the significance of the Chi-square change and the percent correctly predicted (95.2%) Table 15 shows the significance of the predictors in Model One, Block Two.

Table 15

Predictor Statistics for Model One Block Two - Writing

Variable	B	Wald Test	*P*	Exp(B)
Constant	-7.588	62.256	.000	.001
PSAT Verbal	.097	33.280	.000	1.101
PSAT Writing	.179	40.996	.000	1.195
LD	-1.516	31.886	.000	.220
Special Ed	-1.330	17.997	.000	.264
Other				

All of the predictor variables that were added in Block Three were not significant in the model with all other predictors, which was why the researcher chose Block Two. Again, the LD and Special Education Other variables had a negative sign indicating that the chances of a student passing the SOL End-of-Course Writing test decreases when having a special education classification. The equation for this model is:

$$p\hat{} = e^{(-7.692 + .083(\text{PSAT Verbal}) + .182(\text{PSAT Writing}) - .103(\text{Male}) - .047(\text{African-American}) + .328(\text{Asian}) + .765(\text{White}) - 1.471(\text{LD}) - 1.323(\text{SPED Other}))} / 1 + e^{(-7.692 + .083(\text{PSAT Verbal}) + .182(\text{PSAT Writing}) - .103(\text{Male}) - .047(\text{African-American}) + .328(\text{Asian}) + .765(\text{White}) - 1.471(\text{LD}) - 1.323(\text{SPED Other}))}$$

Again, the coefficients (B) are in log-odds units, and the coefficients give the researcher an indication of the relationship between the predictor variables and the dependent variable. The log-odds coefficients indicate how much the predict log-odds of the dependent variable will increase or decrease, depending on the sign of the coefficient.

The odds ratios are shown in the Exp(B) column for easier interpretation. The predicted variable was in terms of a probability of passing the SOL End-of-Course Writing test.

The researcher also examined the classification table for the model, which showed how many cases were correctly predicted. In block two, 95 percent of the cases were correctly predicted, indicating a good fit. However, five percent of cases were incorrectly predicted. The researcher again examined the studentized residuals over two to detect outliers. All of the cases with studentized residuals over two were misclassified as passing the SOL Writing test, when they should have been classified as failing the SOL Writing test. The results were the same for the Writing test as for the Reading test. Most were male, and approximately one-half were White and one-half were African-American. A very small amount of the cases were special education students.

CHAPTER 5

DISCUSSION

Introduction

Chapter 5 begins with a summary of the results from Chapter 4, a comprehensive discussion of the findings, followed by the implications of the study. The limitations of the study follow this discussion. Finally, suggestions for future research are presented.

Summary of Results

This chapter is designed to provide discussion of the results reported in the previous chapter, and explore possibilities for future research. The four research questions posed in this study were:

1. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict the SOL End-of-Course Reading scores?
2. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict the SOL End-of-Course Writing scores?
3. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict whether or not a student will pass the SOL End-of-Course Reading test?

4. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict whether or not a student will pass the SOL End-of-Course Writing test?

The previous chapter shows that the PSAT Verbal and Writing scores contribute the most to the variability in the SOL end-of-course English and Writing scores compared to the gender, race, and special education variables. This is true in both the linear and logistic regression analyses. The PSAT Verbal and Writing scores explain 53 percent of the variance in the SOL Reading scores and 55 percent in the SOL Writing scores. For the linear regression predicting the SOL Reading scores, PSAT Verbal and Writing scores, the No Special Education, and race (African-American) variables were chosen for the equation. In reference to the SOL Writing scores, the equation was very similar, with the addition of the Male variable. The logistic regression analyses included the same variables, with other variables added. For example, the LD and Other Special Education variables contributed to both of the models.

The PSAT Verbal and Writing scores were positively correlated with the SOL Reading and Writing scores. Each time the PSAT scores increase, the SOL scores are predicted to increase as well. If a student is not classified as a special education student, the SOL score was positively impacted. On the other hand, if a student is African-American, the SOL scores on both the Reading and Writing tests are predicted to decrease. If a student is a male, his score on the SOL Writing test is also predicted to decrease.

The logistic regression analyses predict the probability of a student earning a passing score on the SOL Reading and Writing tests. Again, the PSAT Verbal and Writing scores are positively correlated; therefore, the higher the score on the PSAT, the more likely the student is predicted to pass the SOL tests. For the SOL Reading test, the special education indicator variables are negatively correlated with the prediction. The special education variables are also negatively correlated with the probability of passing the SOL Writing test.

Research Question 1

1. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict the SOL End-of-Course Reading scores?

The predictor variables contributed a great deal to predicting the SOL End-of-Course Reading scores. The PSAT/NMSQT Verbal and Writing scores accounted for most of the variance in the predicted scores, with the remaining demographic variables contributing a small amount. However, the special education classification is noteworthy in all of the analyses in this study, and this will be discussed in more detail later in this chapter. In this analysis, all of the race coefficients were positive meaning that SOL Reading scores are predicted to increase in comparison to the referent variable, Other.

Research Question 2

2. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict the SOL End-of-Course Writing scores?

The analysis used to answer this question yielded results similar to the previous question. However, the race (African-American) indicator variable coefficient was negative, but not significant when taking into account all other variables in the model. Again, the special education classifications also had negative coefficients, regardless of being able to utilize accommodations. The PSAT/NMSQT Verbal and Writing scores contributed the most to the chosen model.

The analyses for the first two research questions were separated purposely. The researcher expected the SOL Reading and Writing scores to be positively correlated, but the scores are reported separately. Combining the two SOL scores was a possibility in this research; however this researcher rejected that combination because a student can score higher on one test than the other.

Research Question 3

3. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict whether or not a student will pass the SOL End-of-Course Reading test?

All of the variables contributed to predicting whether or not a student would pass his or her SOL End-of-Course Reading test, with the PSAT/NMSQT Verbal and Writing scores having the most influence. According to the Hosmer-Lemshow Goodness-of-Fit

test, the models did a good job of predicting the outcome. The researcher considered many things in determining the best model. While the percent correctly predicted did not increase a great deal with each block in the model, the difference in the Chi-square value was significantly different between models.

Research Question 4

4. To what extent do the PSAT/NMSQT Verbal and writing scores, gender, race, and special education classification variables predict whether or not a student will pass the SOL End-of-Course Writing test?

The findings here were very similar to the findings in question 3. However, one noteworthy finding was that there was a significant difference between male and female scores on the SOL End-of-Course Writing test. There was no significant difference between male and female scores on the Reading test. The researcher added the gender and race variables in the third block, and the third block did not significantly add to the model. Because the gender and race variables did not add to the model, the researcher chose the second block, which included the PSAT Verbal and Writing scores, and the special education classification. The findings indicate that special education students need more help than non-special education students in writing, which is something that could be used in designing future programs. Again, the Hosmer-Lemshow Goodness-of-Fit test was not significant, which indicated good prediction when comparing the predicted pass rates to the observed pass rates.

Discussion of Results

The results of this study indicate that the PSAT Verbal and Writing scores are good predictors of SOL Reading and Writing test scores, but they are not excellent predictors because they only explain just over 50% of the variance in the dependent variables. Interestingly, the race (African-American) variable has a negative correlation with SOL Reading scores; however, this same variable is positively related to the probability that a student will pass the SOL Reading test. This indicates that African-American students may score lower than their White counterparts on the SOL Reading test; however, they do not score low enough to be predicted to fail the test, which may also signify unequal pass/fail sample sizes. After noting the possibility of unequal pass/fail sample sizes, the researcher re-examined the frequencies of the pass/fail data. For the SOL End-of-Course Reading test, 14% of African-American students failed the test, compared to 3.5% of White students. The SOL End-of-Course Writing test had the same results. In finding that African-American students score lower than White students on the PSAT and SOL tests, this study confirms the findings that Garcia and Fleming (1998) revealed in their study. They examined the correlations between SAT scores and GPA, while this study focused on PSAT scores and SOL scores (which contribute to a student's GPA). Even though they had a small number of cases for several of the categories in their study, their findings were similar to this study.

In another study conducted by Cahalan, Mandinach, and Camara (2002), the researchers provided predictive validity evidence of the SAT. They stated that the test scores can be used with a student's GPA to predict freshman year (college) GPA,

particularly for students with learning disabilities. The researchers found that a student with a disability scores lower than students without a disability. This finding is supported in the current research. In the linear regression analyses with SOL Reading score as the dependent variable, a student with no special education classification had a regression coefficient of 17.203. When predicting the SOL Reading test score, this means that a student who is not classified as special education is predicted to score 17.203 points higher than a special education student. Results are similar when predicting the SOL Writing scores. In this case, the regression coefficient for students who do not have any special education classification is 25.907. This finding is also reflected in the logistic regression analyses.

Finally, the findings in this study were very similar to the findings in the Warry (2003) study, which used the PSAT Verbal and Writing scores to predict performance on the Massachusetts statewide standards-based test, the Massachusetts Comprehensive Assessment System (MCAS). Warry excluded race and gender from her equations because they contributed the least to the predictive ability. Warry used PSAT Verbal and Writing scores; however, the PSAT Writing scores are different from the PSAT Writing scores in this study. At the time of Warry's study, the writing portion of the PSAT was only multiple choice. During the current study, the writing portion was the newly developed composition component. While this study kept race and gender in the equations, these variables did not contribute as much to the prediction as did the PSAT Verbal and Writing scores, just as in Warry's study. Warry's study also included a community (urban or suburban) variable. Including a community variable was not

possible in this research; however, an interesting future possibility would be to identify and include a variable that denotes that the student attends a school with a certain percentage of students on free and/or reduced lunch.

Because the PSAT Verbal and Writing scores were highly correlated with the SOL Reading and Writing scores, PSAT scores could be used to identify students at risk of not passing their English end-of-course test, which is just one of the graduation requirements in Virginia. The students are given multiple chances to re-take the SOL test in order to pass; however, if the schools can use a tool such as this, it can only benefit students, teachers, and entire school systems by decreasing the number of multiple administrations. Additionally, the residual and classification results would be even more helpful to teachers. The classification results from the logistic regression analyses could be used to determine which types of students were correctly predicted to pass the SOL End-of-Course Reading and Writing tests, such as special education students. The particular school system that was used in this study has a wide variety of schools, with a diverse student population. Other school districts with a similar population could use the results in helping to identify students at risk of not passing their statewide standards-based tests. The researcher cautions other school districts to test the model with their population prior to making any determinations about particular students. Finally, In Snyder's (2004) study, the researcher found that failing an SOL End-of-Course test (not necessarily Reading or Writing) impacted students' decision to drop out of school prior to graduation. If the PSAT/NMSQT Verbal and Writing scores could be used, even

minimally to help student pass their SOL tests, it could perhaps prevent a few at-risk students from dropping out prior to graduation

The *No Child Left Behind Act of 2001* requires that all students pass their statewide tests by 2014. If the PSAT scores can be used to help identify students for remediation, then the cost of the PSAT testing would be worth the gains of helping students pass the SOL End-of-Course Reading and Writing tests. In addition, NCLB requires that schools show Adequate Yearly Progress, meaning that each school has to show that test scores have improved from year to year. In states that have high stakes for schools, the PSAT scores could be an invaluable resource for detecting students' areas of improvement, especially since the PSAT score reports delineate students' areas of strength and weakness. Students who score well on the PSAT tests could be utilized as resources for others who need additional assistance, thus reducing the need for multiple administrations of the English SOL tests to particular students. The additional preparation devoted to these lower performing students could provide the confidence and knowledge that they need to pass the SOL English test the first time they take it.

Practical Implications, Future Research, and Conclusion

Limitations of the Study

The first limitation of this study is that the PSAT test scores have not been validated to be used as predictors for the SOL tests. However, this study provides evidence to support this additional predictive validity of the PSAT Verbal and Writing tests. In addition, the PSAT is an aptitude test, which does not necessarily measure the

same constructs as the SOL tests, which are standards-based tests. This limitation can also be argued against because the College Board (2004c) aligned the PSAT and SAT items with the Virginia English and Math Standards of Learning (See Appendix for aligned standards).

Another limitation is that student GPA and socio-economic status were not used in the study. Most of the studies which use SAT scores to predict college freshman GPA also use high school GPA in the prediction. The researchers discussed in the review of literature even state that SAT scores, along with high school GPA are the best predictors of a student's academic performance during the first year of college (Camara, 2000; Frisch-Kowalski, 2003; Frontline, 1999). Camara and Millsap (1998) did use the PSAT scores to predict scores on Advanced Placement tests. The AP tests are subject area tests, just like the SOL tests. They found that a strong correlation exists between PSAT and AP scores, and that this relationship is actually stronger than those between GPA and AP test scores. Student GPA could have added to this study since other research has found that the GPA variable improves prediction of college GPA when added to the equation. In all statistical analyses, the possibility exists that the researcher is not collecting data on other variables that could be affecting the outcomes. Variables denoting GPA and SES could have added to the regression models.

One final limitation is that standards-based tests are not expected to be normally distributed. Because the current state-wide accountability tests are measuring students' knowledge of the basic standards, students are expected to pass them, which can positively skew the distribution. This data set did have a high number of perfect scores on

the SOL Reading and Writing tests, meaning that the data were not normally distributed; however, outliers were examined and considered in the analyses.

Practical Implications

The current study has contributed to the scientifically-based research that is so crucial to the *No Child Left Behind Act of 2001*. Scientific evidence is crucial to making improvements in education, which is a cornerstone to *No Child Left Behind.* In reference to the necessary principles for scientific evidence, the researcher was able to use theory and other research to develop the methodology for the study. In addition, the methodology allowed for directly investigating the research questions. This research has provided support showing that the PSAT/NMSQT Verbal and Writing scores can be used to help in predicting the SOL End-of-Course Reading and Writing scores.

The PSAT/NMSQT scores have been shown to predict AP test scores, which are also substitute tests for the SOLs. However, this study shows that the PSAT/NMSQT no longer needs to be used just for the college-bound students in the advanced classes. Students who may be in jeopardy of not graduating from high school may be able to receive additional instruction to help them meet the required basic standards. This is particularly true for special education students who may not consider certain colleges (or college at all) because their standardized and standards-based test scores are lower than their non-special education counterparts. Findings in this study support the findings in Cahalan, Mandinach, and Camara's (2002) study on the predictive validity of the SAT test, particularly with students with learning disabilities. Regardless of the

accommodations designed to help those students, they still scored significantly lower than students with no disability. Support programs for special education students are in practice at this time; however, the students still yield lower scores. Scientifically-based research needs to be utilized to see if those programs are working. The Educational Testing Service provides specific feedback to each student, outlining how they can improve scores on future PSAT/NMSQT and SAT tests. This feedback can also be used in conjunction with the additional instruction for any student.

The PSAT/NMSQT could also be administered earlier. The school district used in this study administers the PSAT/NMSQT during the students' junior year. The students could take the PSAT/NMSQT during their freshman or sophomore year, which would allow more time for remediation and preparation for the upcoming end-of-course tests. Additionally, many of the other end-of-course tests in Virginia are taken during the freshman or sophomore year, such as Algebra. If the PSAT/NMSQT math scores could be validated and used for the purpose of preparing students for math end-of-course tests, this would definitely be support for earlier administration of the PSAT/NMSQT.

Future Research

This study could be replicated in other school districts, particularly those with high stakes, standards-based tests. In fact, the Warry study, which found similar results, was completed in school districts with high stakes tests. The GPA variable and other variables to indicate socio-economic status, percentage of students receiving free and reduced lunch or school size might add to the prediction capability of the regression

models, which could lead to a higher prediction of the variance in the predictor variable. In the equations in this study, over 40% of the variance was not explained.

Other studies could also be developed using the PSAT Math scores. In fact, an interesting study would be to add the PSAT Math scores to this study, which might indicate that the math tests depend on a student's ability to read as much as the English tests do. The PSAT Math scores could also be used to predict scores on statewide math end-of-course tests.

The study could also be replicated in school districts that have high Hispanic populations. The Hispanic population is the fastest growing minority population in the United States, which means that the number of Hispanic students taking the PSAT/NMSQT is growing as well. The literature revealed evidence that Hispanic students score lower on standardized tests than White students. Gandara and Lopez (1998) found that Hispanic students did not score as well on standardized tests as White students, and that low scores result in low self-esteem for the Hispanic students. The regression equations in this study could also be validated with new data. In addition, other demographic variables could be used, such as those in the Warry (2003) study. Additionally, the model should be tested on a new group. Finally, the categories of failing, pass proficient, and pass advanced on the SOL End-of-Course tests could be examined in a logistic regression.

Conclusion

The research on the use of standardized testing has brought both criticism and support. Evidence supporting validity and reliability is sufficient for both the PSAT/NMSQT and the Virginia SOL tests, yet other researchers have found evidence that does not support the validity and reliability of the test scores. In particular, colleges have started to question the SAT as an admission requirement. The president at the University of California, Richard Atkinson, stated that the test score gap between African-American students and White students is smaller on the SAT II Subject tests than on the SAT test (Schrag, 2001a). Atkinson stated that his data show that the SAT test scores are not valid or reliable. As for the SOLs, the literature supports an accountability system in our public schools, yet criticizes the high stakes attached. The juxtaposition of findings in the research perpetuates the debate that focuses on these tests.

The specific conclusions resulting from this study are:

- Special education students are predicted to score lower on the SOL End-of-Course Reading and Writing tests.
- The PSAT/NMSQT Verbal and Writing scores account for most of the variability in the SOL End-of-Course Reading and Writing scores.
- Male and African-American students are predicted to score lower on the SOL Writing test than other students, excluding special education.
- The PSAT/NMSQT Verbal and Writing test scores are good predictors of the SOL End-of-Course Reading scores.

- The PSAT/NMSQT Verbal and Writing test scores are good predictors of the SOL End-of-Course Writing scores
- The PSAT/NMSQT Verbal and Writing test scores are good predictors of whether or not a student will pass or fail the SOL End-of-Course Reading test.
- The PSAT/NMSQT Verbal and Writing test scores are good predictors of whether or not a student will pass or fail the SOL End-of-Course Reading test.

This study has also contributed to the scientifically-based research literature. The research followed the six principles as outlined in *Scientific Research in Education* (Shavelson & Towne, 2002). First, the research questions in this study were conducive to empirical examination, and the research is grounded in applicable theory. Next, the regression analyses used allowed the researcher to directly answer the posed questions. The reasoning used in this study was logical and specific. Finally, this study could be replicated, the results are generalizable, and this study open to critical discourse.

This study has contributed to the existing literature by providing scientific evidence for using a well-known test for diagnostic purposes; for using the PSAT/NMSQT scores to evaluate more than just the advanced, college-bound students; and for generalizing results to school districts with a similar population. In addition, multiple methods have been utilized to collect evidence supporting validity and reliability of scores for all of the tests used in this study. This study provides two additional methods (linear regression and logistic regression). Finally, because the PSAT/NMSQT

Writing test is so new, college and university admissions offices have not yet determined how they will use these particular test scores. This study allows for immediate use of these brand new test scores.

LIST OF REFERENCES

List of References

Albemarle County Schools (2005). *2005 PSAT achievement test results.* Retrieved March 31, 2006, from http://schoolcenter.k12albemarle.org/education/sctemp/490232ab68a4d12e0ca2d89d0696b1cb/05PSAT.pdf

American College Testing Program (2004). *ACT Newsroom: Facts about the ACT assessment*. Retrieved October 18, 2004, from http://www.act.org/news/aapfacts.html

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Breland, H., Kubota, M., Nickerson, K., Trapani, C., & Walker, M. (2004). *New SAT writing prompt study: Analyses of group impact and reliability* (College Board Research Report No. 2004-1, ETS RR-04-03). New York: College Board.

Cahalan, C., Mandinach, E., & Camara, W. (2002). *Predictive validity of SAT I: Reasoning Test for test takers with learning disabilities and extended time accommodations*. New York: College Board.

Camara, W., & Echternacht, G. (2000). The SAT I and high school grades: Utility in predicting success in college. *The College Board, Research Notes*. The College Board, Office of Research and Development.

Camara, W., & Millsap, R. (1998). *Using the PSAT/NMSQT and course grades in predicting success in the Advanced Placement Program.* New York: College Board Publications.

Chandler, M. (Executive Producer). (1999, October 5). *Frontline: Secrets of the SAT* [Television broadcast]. New York and Washington, DC: Public Broadcasting Service.

Crouch, K. (2003). The relationship of student performance on mathematics and reading exit-level tests. *Dissertation Abstracts International, 64* (12A), 4283. (UMI No. 3118102)

Data and Statistical Services (2006). *Introduction to regression*. Retrieved March 31, 2006 from http://dss.princeton.edu/online_help/analysis/regression_intro.htm

Dowlin, W. (Winter, 2000). Enemies of promise: Why America needs the SAT. *Academic Questions, 13*, 6.

Education Commission of the States (2006). *Assessment: National tests*. Retrieved January 6, 2006 from http://www.ecs.org/html/issue.asp?issueid=12&subIssueID=77

Education Week (2006). *Quality Counts 2006.* Retrieved January 5, 2006 from http://www.edweek.org/qc06

Freedman, M. (2003). Disabling the SAT: how the College Board is undermining its premier test. *Education Next, 3,* 36-43.

Frisch-Kowalski, S. (2003). *The SAT: A timeline of changes*. New York: College Entrance Exam Board.

Gandara, P., & Lopez, E. (1998). Latino students and college entrance exams: How much do they really matter? *Hispanic Journal of Behavioral Sciences, 20,* 17-38.

Garcia, N., & Fleming, J. (1998). Are standardized tests fair to African Americans? *Journal of Higher Education, 69*, 471-475.

Greene, J., Winters, M., & Forster, G. (2003). *Testing high stakes tests: Can we believe the results of accountability tests?* New York: Manhattan Institute, Center for Civic Education. (ERIC Document Reproduction Service No. ED475488)

Haney, W., Madaus, G., & Lyons, R. (1993). *The fractured marketplace for standardized testing.* Norwell, MA: Kluwer Academic Publishers.

Hilliard III, A. (2000). Excellence in education versus high-stakes standardized testing. *Journal of Teacher Education, 51,* 293.

Kohn, A. (2000). Burnt at the high stakes. *Journal of Teacher Education, 51,* 315.

Moon, T., Brighton, C., & Callahan, C. (2003). State standardized testing programs: Friend of foe of gifted education? *Roeper Review, 25,* 49-50.

Myers, R. (1990). *Classical and modern regression with applications*. Pacific Grove, CA: Thomson Learning.

National Center for Education Statistics (2004). *The nation's report card: State profile: VA.* Retrieved May 10, 2005 from http://nces.ed.gov/nationsreportcard/states/profile.asp?state=VA

National Conference of State Legislators (2005). *No child left behind: History.* Retrieved March 14, 2006 from http://www.ncsl.org/programs/educ/NCLBHistory.htm

North Central Regional Educational Laboratory (2004). *Standardized tests.* Retrived March 14, 2006 from http://www.ncrel.org/sdrs/areas/issues/students/earlycld/ea5lk3.htm

Office of Educational Research and Improvement (2002). *Office of educational research and improvement.* Retrieved March 11, 2006 from http://www.ed.gov/offices/OERI/Index.html

Organ, D. (May 2001). Predicting success in school and at work. *Business Horizons, 44*, 1.

Palin, R. (2001). PSAT and AP Success. *OAH Magazine of History, 15*(3), 55-56.

Porter, K. (2002). *The value of a college degree.* Washington, DC. (ERIC Document Reproduction Service No. ED470038)

Sacks, P. (1997). *Standardized testing: Meritocracy's crooked yardstick, 29*, 24-31.

Schrag, P. (2001a). Going holistics: Scrapping the SAT I test. *The American Prospect, 12*, 18.

Schrag, P. (2001b). War on the SAT. *The American Prospect, 13*, 24-27.

Schuppan, Jr. F., Curley, W., O'Callaghan, R., & Schmidt, A. (2004, April). *Content changes to the current SAT I: Reasoning Test.* Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

Shavelson, R., & Towne, L. (Eds.). (2002). *Scientific Research in Education.* Washington, D.C.: National Academy Press.

Sireci, S. (2004). The SAT. *Education Next, 4*, 5-6.

Sireci, S. (2005). Unlabeling the Disabled: A Perspective on Flagging Scores from Accommodated Test Administrations. *Educational Researcher, 34, 3-12.*

Snyder, A. (2004). The impact of high stakes tests on high school students' decisions to drop out of school (Doctoral dissertation, Virginia Commonwealth University, 2004). *Dissertation Abstracts International, 66,* 157.

Technical Advisory Committee. (2001). *Review of selected technical characteristics of the Virginia Standards of Learning (SOL) assessments.* Retrieved November 2, 2004, from http://www.pen.k12.va.us/VDOE/Assessment/virginiareport.pdf

The College Board (2004a). *The New SAT 2005.* Retrieved September 30, 2004, from http://www.collegeboard.com/newsat/index.html

The College Board (2004b). *SAT reasoning test.* Retrieved May 10, 2005 from http://www.sat.org/prod_downloads/counselors/hs/sat/resources/handbook/2_SAT%20Program.pdf

The College Board (2004c). *Virginia curriculum standards and areas of alignment PSAT/NMSQT, SAT I and SAT II: Writing.* Retrieved May 10, 2005 from http://www.pen.k12.va.us/VDOE/Instruction/AlignRpt_VA.pdf

The College Board (2006a). *ACES validity handbook.* Retrieved March 15, 2006 from http://collegeboard.com/highered/apr/aces/vhandbook/evidence.html

The College Board (2006b). *Frequently asked questions.* Retrieved March 14, 2006 from http://www.collegeboard.com/student/testing/sat/about/sat/FAQ.html

Trends in International Mathematics and Science Study (2006). *Frequently asked questions about the assessment.* Retrieved March 23, 2006 from http://nces.ed.gov/timss/FAQ.asp?FAQType=1

U.S. Department of Education (2001). *Stronger Accountability.* The No Child Left Behind Act of 2001. Washington, DC: U.S. Department of Education.

U.S. Department of Education (2002). *U.S. Department of Education Awards Contract for "What Works Clearinghouse."* Retrieved March 11, 2006 from http://www.ed.gov/news/pressreleases/2002/08/08072002a.html

Virginia Department of Education Division of Assessment and Reporting. (1999, February). *Standards of learning (SOL) tests: Validity and reliability information spring 1998 administration.* Retrieved February 20, 2006 from http://www.pen.k12.va.us/VDOE/Assessment/validity.PDF

Virginia Department of Education (1999, July 29). *Virginia students show improvement on all SOL tests in second round.* Retrieved September 30, 2004, from http://www.pen.k12.va.us/VDOE/NewHome/pressreleases/jul2999.html

Virginia Department of Education. (2000, October). *Virginia Standards of Learning assessments: Virginia technical report.* (Report). Virginia Department of Education. Richmond, VA: Author.

Virginia Department of Education. (2001). *Every child can succeed: A parent's guide to Virginia's Standards of Learning program.* Richmond, VA: VDOE. (ERIC Document Reproduction Service No. ED481126)

Virginia Department of Education (2002). *Virginia's consolidated state application for state grants under the Title IX, Part C, Section 9302 of the Elementary and Secondary Education Act.* (Public Law 107-110). Richmond, VA: Virginia Department of Education.

Virginia Department of Education (2003, March). *Virginia's NCLB assessment and accountability plan.* Paper presented at the Key Instructional Leadership Forums on NCLB, Richmond, VA.

Virginia Department of Education (2004). *Student membership PK-12.* Retrieved May 10, 2005 from http://www.pen.k12.va.us/VDOE/dbpubs/Fall_Membership/2004/readme.html

Virginia Department of Education (2006a). *2 + 4 in 2004...and 2005 and 2006.* Retrieved January 21, 2006 from http://www.pen.k12.va.us/2plus4in2004/2plus4gradinfo.pdf

Virginia Department of Education (2006b). *National assessment of educational progress: Frequently asked questions*. Retrieved May 23, 2006 from http://www.pen.k12.va.us/VDOE/Assessment/NAEP/FrequentlyAskedQuestions.htm

Warry, J. (2003). An analysis of variables affecting standardized test results at the high school level. *Dissertation Abstracts International, 64* (54A), 1519. (UMI No. 3090418)

Westchester Institute for Human Services Research (April 2003). *High-stakes testing.* Retrieved March 14, 2006 from http://www.sharingsuccess.org/code/bv/testing.pdf

Zucker, Sasha. (2003, December). *Fundamentals of standardized testing.* (Report). San Antonio, TX: Harcourt Assessment, Inc.

APPENDIX

Appendix

Table 1

English SOLs Aligned with PSAT/NMSQT

English Grade Level – 9	
9.3 The student will read and analyze a variety of literature.	
• A) Identify format, text structure, and main idea.	PSAT/NMSQT SAT I: Verbal SAT II: Writing
• C) Use literary terms in describing and analyzing selections.	PSAT/NMSQT SAT I: Verbal
9.4 The student will read and analyze a variety of informational materials (manuals, textbooks, business letters, newspapers, brochures, reports, catalogs) and nonfiction materials including journals, essays, speeches, biographies, and autobiographies.	
• E) Extend general and specialized vocabulary through speaking, reading and writing.	PSAT/NMSQT SAT I: Verbal SAT II: Writing
9.6 The student will develop narrative, expository, and informational writings to inform, explain, analyze, or entertain.	
• F) Arrange paragraphs into a logical progression.	PSAT/NMSQT SAT II: Writing
• H) Proofread and prepare final product for intended audience and purpose.	PSAT/NMSQT
9.7 The student will edit writing for correct use of language, sentence formation, punctuation, capitalization, and spelling as part of the writing process.	
• A) Use and apply rules for the parts of a sentence including: subject/verb, direct/indirect object and predicate nominative/predicate adjective.	PSAT/NMSQT SAT II: Writing

• B) Use parallel structures across sentences and paragraphs.	PSAT/NMSQT SAT II: Writing
• C) Use appositives and main/subordinate clauses.	PSAT/NMSQT SAT II: Writing
• D) Use commas and semicolons to distinguish and divide main and subordinate clauses.	PSAT/NMSQT SAT II: Writing
English Grade Level – 10	
10.3 The student will read, comprehend, and critique literary works.	
• B) Identify main and supporting ideas.	PSAT/NMSQT SAT I: Verbal SAT II: Writing
• C) Make predictions, draw inferences, and connect prior knowledge to support reading comprehension.	PSAT/NMSQT SAT I: Verbal
• D) Explain similarities and differences of techniques and literary forms represented in the literature of different cultures.	PSAT/NMSQT
• E) Identify universal themes prevalent in the literature of all cultures.	PSAT/NMSQT
10.4 The student will read and interpret informational materials.	
• A) Analyze and apply the information contained in warranties, contracts, job descriptions, technical descriptions, and other informational sources, including labels, warnings, manuals, directions, applications, and forms, to complete specific tasks.	PSAT/NMSQT SAT I: Verbal
10.7 The student will develop a variety of writings with an emphasis on exposition.	
• D) Organize ideas into a logical sequence.	PSAT/NMSQT SAT II: Writing

• F) Proofread and prepare final product for intended audience and purpose.	PSAT/NMSQT
10.8 The student will edit writing for correct grammar, capitalization, punctuation, spelling, sentence structure, and paragraphing.	
• B) Apply rules governing use of the colon.	PSAT/NMSQT SAT II: Writing
• C) Distinguish between active voice and passive voice.	PSAT/NMSQT SAT II: Writing
English Grade Level – 11	
11.3 The student will read and analyze relationships among American literature, history, and culture.	
• C) Discuss American literature as it reflects traditional and contemporary themes, motifs, universal characters, and genres.	PSAT/NMSQT
11.4 The student will read and analyze a variety of informational material.	
• A) Use information from texts to clarify or refine understanding of academic concepts.	PSAT/NMSQT SAT I: Verbal
• B) Read and follow directions to complete an application for college admission, for a scholarship, or for employment.	PSAT/NMSQT SAT I: Verbal
• E) Analyze information from a text to draw conclusions.	PSAT/NMSQT SAT I: Verbal
11.6 The student will read and critique a variety of dramatic selections.	
• C) Explain the use of verbal, situational, and dramatic irony.	PSAT/NMSQT SAT I: Verbal
11.7 The student will develop a variety of writings with an emphasis on persuasion.	
• D) Organize ideas into a logical sequence.	PSAT/NMSQT SAT II: Writing

• H) Proofread final copy and prepare document for intended audience or purpose.	PSAT/NMSQT
11.9 The student will write, revise, and edit personal, professional, and informational correspondence to a standard acceptable in the workplace and higher education.	
• C) Present information in a logical manner.	PSAT/NMSQT SAT II: Writing
• E) Use technology to access information, plan a composition, and develop writing.	PSAT/NMSQT
11.10 The student will analyze, evaluate, synthesize, and organize information from a variety of sources to produce a research product.	
• G) Edit writing for clarity of content and effect.	PSAT/NMSQT
• H) Edit copy for grammatically correct use of language, spelling, punctuation, and capitalization.	PSAT/NMSQT
• I) Proofread final copy and prepare for publication or other use.	PSAT/NMSQT
• J) Use technology to access information, organize ideas, and develop writing.	PSAT/NMSQT
English Grade Level – 12	
12.3 The student will read and analyze the development of British literature and literature of other cultures.	
• C) Relate literary works and authors to major themes and issues of their eras.	PSAT/NMSQT
12.4 The student will read and analyze a variety of informational materials, including electronic resources.	
• A) Identify formats common to new publications and information resources.	PSAT/NMSQT SAT I: Verbal SAT II: Writing
• B) Recognize and apply specialized informational vocabulary.	PSAT/NMSQT

	SAT I: Verbal SAT II: Writing
12.5 The student will read and critique a variety of poetry.	
• A) Explain how the choice of words in a poem creates tone and voice.	PSAT/NMSQT SAT I: Verbal
• B) Explain how the sound of a poem (rhyme, rhythm, onomatopoeia, repetition, alliteration, assonance, and parallelism) supports the subject and mood.	PSAT/NMSQT SAT I: Verbal
12.7 The student will develop expository and informational writings.	
• F) Apply grammatical conventions to edit writing for correct use of language, spelling, punctuation, and capitalization.	PSAT/NMSQT
• G) Proofread final copy and prepare document for publication or other use.	PSAT/NMSQT
12.8 The student will write documented research papers.	
• F) Edit copies for correct use of language, capitalization, punctuation, and spelling in final copies.	PSAT/NMSQT
• G) Proofread final copy and prepare document for publication or other use.	PSAT/NMSQT

Vita

Susan P. McKelvey was born on September 14, 1971 in Richmond, Virginia. She graduated from Mills E. Godwin High School in Richmond, Virginia in 1989. She received a B.A. in French from Randolph-Macon College in Ashland, Virginia in 1993. She completed her teacher certification at Virginia Commonwealth University in 1996 and subsequently taught French for two years at Louisa County High School. She received a Master of Education in Adult Education from Virginia Commonwealth University in 2002.